



Invasion du nématode du pin *Bursaphelenchus xylophilus*

Sophie Mallez

► To cite this version:

Sophie Mallez. Invasion du nématode du pin *Bursaphelenchus xylophilus*. Sciences agricoles. Université Nice Sophia Antipolis, 2014. Français. NNT : 2014NICE4099 . tel-01127300

HAL Id: tel-01127300

<https://theses.hal.science/tel-01127300>

Submitted on 7 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nice Sophia-Antipolis

UFR Sciences

Ecole Doctorale des Sciences de la Vie et de la Santé

THÈSE

préparée au sein de l'Institut Sophia-Agrobiotech, UMR INRA 1355/UNS/CNRS 7254
présentée publiquement à l'INRA de Sophia-Antipolis

pour obtenir le titre de **Docteur en Sciences** de l'Université de Nice Sophia-Antipolis

Discipline : Biologie

Mention : Biologie des interactions et écologie

INVASION DU NÉMATODE DU PIN

Bursaphelenchus xylophilus

par

Sophie Mallez

Soutenue le 2 Décembre 2014 devant le jury composé de :

Géraldine Roux-Morabito	Maitre de Conférences, Université d'Orléans	Examinatrice
Isabel Abrantes	Professeur, Université de Coimbra	Examinatrice
Arnaud Estoup	Directeur de Recherches, INRA Montpellier	Rapporteur
Olivier Plantard	Chargé de Recherches, INRA Nantes	Rapporteur
Philippe Castagnone	Directeur de Recherches, INRA Sophia-Antipolis	Directeur de thèse
Thomas Guillemaud	Directeur de Recherches, INRA Sophia-Antipolis	Co-Encadrant

Résumé

Les invasions biologiques se définissent comme la prolifération d'organismes dans de nouvelles aires géographiques souvent distantes de leur aire originelle. Ce sont des phénomènes essentiellement irréversibles et leur nombre est croissant. La prolifération se fait bien souvent au détriment des espèces natives et engendre des conséquences écologiques et économiques importantes. Réussir à comprendre les processus qui conduisent au succès d'une invasion est donc crucial.

Dans ce projet, nous nous sommes intéressés à l'invasion du nématode du pin, *Bursaphelenchus xylophilus*, un ver microscopique responsable de la maladie du dépérissement du pin et causant la perte de millions d'arbres annuellement à travers le monde. Notre objectif global était d'affiner nos connaissances sur ce cas d'invasion et particulièrement d'établir l'histoire d'invasion des populations européennes du nématode du pin.

Après avoir développé les marqueurs microsatellites nécessaires à notre étude, nous avons analysé la diversité génétique neutre d'échantillons de populations naturelles de nématode du pin de l'aire native et de différentes zones envahies à l'aide de méthodes de génétique des populations. Ces analyses ont essentiellement mis en évidence que (i) l'aire native du nématode du pin est fortement structurée à fines échelles spatiales et présente une diversité génétique faible à modérée, suggérant un rôle important de la dérive génétique, non compensée par la dispersion, (ii) les populations européennes présentent une diversité génétique extrêmement faible, suggérant un seul évènement d'introduction en Europe, et que (iii) l'Amérique du Nord est la source la plus probable des populations envahissantes européennes, non le Japon et la Chine. Une exploration préliminaire des raisons du succès a aussi mis en évidence des différences en terme de vitesse de croissance entre individus issus

de populations naturelles native et envahissantes. Il est nécessaire de conduire de nouvelles analyses pour avoir une vision claire des facteurs qui ont pu conditionner le succès de l'invasion du nématode du pin. Ce projet a également souligné l'importance de travailler avec des échantillons de populations naturelles, plutôt que des souches de collection/laboratoire, pour réaliser des inférences sur l'histoire des populations et leurs relations.

Ce projet souligne l'intérêt que les nématologistes et les généticiens des populations ont à collaborer. La génétique des populations a en effet permis de clarifier l'histoire de l'invasion des populations européennes et ce modèle nématode, du fait de son cycle de vie complexe résultant en un cas extrême de faible diversité génétique, a permis de s'interroger sur des méthodes et analyses couramment utilisées en génétique des populations.

Abstract

Biological invasions are the proliferation of organisms in new areas geographically distant from their origin. These are irreversible phenomena and their number is growing. Such proliferation is often detrimental to native species and cause important economical and ecological consequences. Improving our understanding of the process(es) leading to successful invasions is therefore crucial.

In this project, we were interested in the invasion of the pinewood nematode, *Bursaphelenchus xylophilus*, a microscopic worm responsible for the pine wilt disease and causing the annual loss of millions of trees worldwide. Our global objective was to improve our knowledge on this case of invasion and particularly to clarify the invasion history of European populations.

After developing the microsatellite markers needed for our study, we analysed the genetic diversity of samples of natural populations from both the native area and invaded areas thanks to population genetics methods. These analyses revealed that (i) the native area is strongly genetically structured at fine scales and displays a low to moderate genetic diversity, suggesting an important role of the genetic drift, not compensated by dispersal, (ii) the European populations exhibit an extremely low genetic diversity, indicating a unique event of introduction in Europe and (iii) North America is the most probable source of European populations compared to China and Japan. A preliminary investigation of the causes of the invasion success of the pinewood nematode also revealed significant differences in growth between native and invasive individuals of natural populations. New analyses are necessary to obtain a precise view concerning the factors that may contribute to the success of invasion. As a complement to the initial objectives, this project also highlights the importance to work with

samples from natural populations, rather than laboratory/collection strains, to explore the populations' history and their relationships.

This project is finally a good example of the benefits of collaboration between nematologists and population geneticists. Population genetics allowed to clarify the invasion history of European populations and this “nematode” model, because of its complex life cycle resulting in an extreme case of low genetic diversity, provided the opportunity to question methodologies in population genetics that are routinely used.

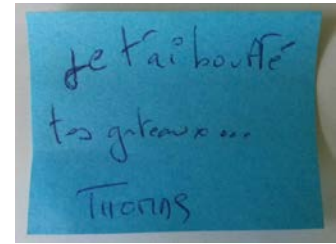
Remerciements

Je souhaite tout d'abord remercier les membres du jury, Géraldine Roux-Morabito, Isabel Abrantes, Arnaud Estoup et Olivier Plantard d'avoir accepté mon invitation.

Je souhaite ensuite remercier chaleureusement mes deux directeurs de thèse, Thomas et Philippe. Un grand merci à tous les deux pour m'avoir fait confiance sur ce projet. Ce fut vraiment une très bonne expérience pour moi et j'ai beaucoup appris pendant ces trois années à vos côtés. Je vous remercie pour votre bonne humeur, votre disponibilité et vos précieux conseils. Je n'ai manqué de rien et j'ai bien conscience d'avoir eu beaucoup de chance de voyager autant ! D'un naturel stressé, ces trois années ont été l'occasion de nombreux doutes et vous avez su me rassurer et me soutenir, alors un grand merci pour ça aussi ! Particulièrement Thomas... tu écrivais dans tes propres remerciements : "Ces deux chefs sont essentiellement disponibles, peut-être trop. Car les habitudes qu'ils ont déteignent



et s'héritent. Donc danger."... Quelle intuition tu as eu ! L'homme que l'on pourrait appeler le Lucky Luke de la relecture, je te remercie pour ta réactivité dans nos échanges... qui t'a peut-être provoquée des petites fringales... !! Particulièrement Philippe... je te remercie de m'avoir fait découvrir le monde mystérieux de la nématologie et de ne pas avoir hésité à mettre à contribution tes muscles pour la bonne réussite de ce projet... !! Enfin voilà, je pourrais en écrire des pages alors je finirai simplement en vous disant que si c'était à refaire, je ne changerai certainement pas l'encadrement que j'ai reçu. Je le considère de qualité et je souhaite le même à tous les thésards.



Chantalinox, mon petit rayon de soleil et mon acolyte de la première heure ! Je te remercie de m'avoir accompagnée dans toutes les aventures de cette thèse et dans la découverte des nématodes. Tu as toujours été de précieux conseils pour dompter ces petites bêtes ! Ces trois années n'ont pas toujours été facile pour moi et tu as toujours été là... merci pour ça. Merci aussi de m'avoir transmis la compétence improbable de pêche de nématode à la moustache de chat ! Je voudrais juste te dire une dernière chose...ce n'était carrément pas moisi de travailler avec toi et on remet ça quand tu veux !!

Un grand merci à toute l'équipe BPI pour son accueil et pour les supers moments partagés... Merci pour les pots, les sorties et les escapades BPI. J'ai des souvenirs plein la tête grâce à vous ! Merci donc à Éric, Didier, Alexandra, Elodie, Thibaut, Ferran, Aurélie, Xavier, Isabelle, Isabelle, Émeline, Géraldine, Claire et toutes les personnes qui sont passées dans l'équipe. Je pense notamment à Virginie et Maud. Une pensée spéciale pour Éric... Merci d'avoir pallié mes lacunes musicales en me faisant découvrir des classiques et aussi...des moins classiques !! Merci aussi d'avoir partagé avec moi tes connaissances sur l'ABC.

Un grand merci également à l'équipe IPN que je connais moins, je dois l'avouer. Merci pour votre accueil et pour m'avoir fait découvrir une spécialité locale : le pain bagnat ! Un merci tout particulier au labo "Casta", je pense notamment à Karine et Nathalie. Merci pour

les petites pauses gourmandes et les fous rires, qui ont rythmé mes petites visites hebdomadaires voire quotidiennes.

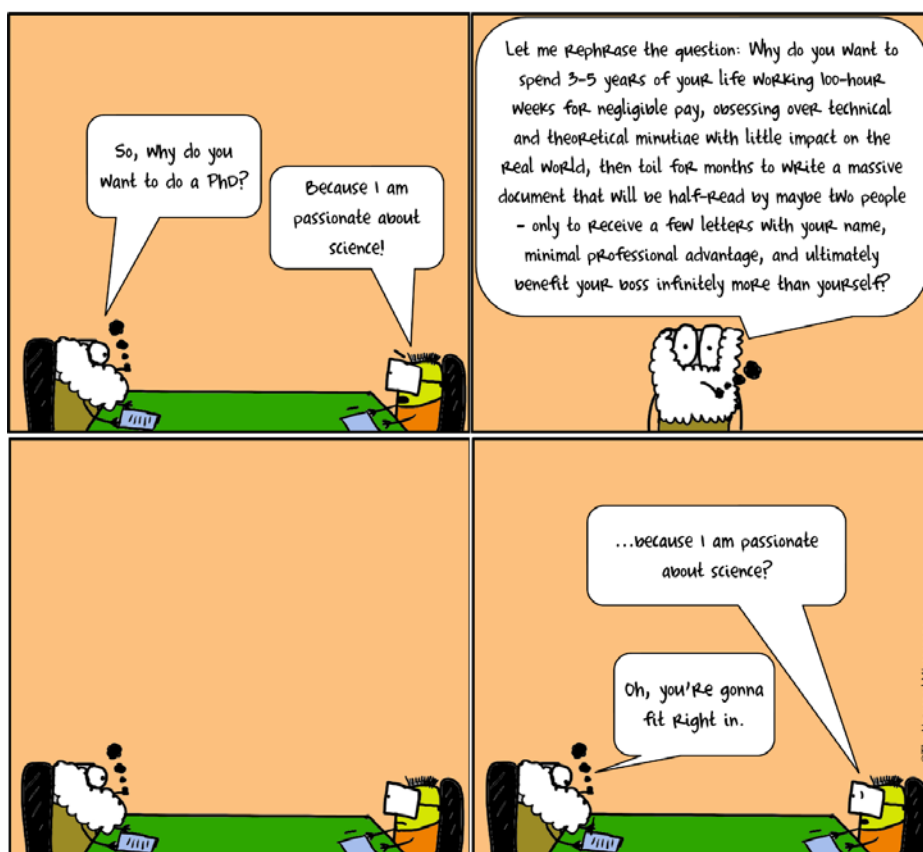
Un immense merci aux thésards du bureau : Thibaut, Margarita, Gérald, Chloé et Bastien. Merci à eux pour la super ambiance du bureau, pour les soirées, pour nos rêves farfelus qui m'ont bien fait rire et pour avoir partagé mes joies et mes doutes. Merci à Margarita pour ces gâteaux toujours délicieux. Merci à Thibaut pour ses conseils et son écoute. Merci à Gérald pour l'opportunité qu'il m'a donné de faire du terrain, j'ai adoré.

Merci à Arnaud...on a commencé quasiment en même temps et on finit quasiment en même temps. C'était super de partager les grands moments ensemble ! Merci pour les échanges d'infos, le soutien, les moments de détente et j'en passe ! Merci aussi à Hugo pour ces jeux de mots et pour nos petites discussions...ponctuelles mais toujours très intéressantes !

Je remercie également toutes les personnes qui ont contribué de près ou de loin à ce projet. Je pense notamment à Margarida Espada, Manuel Mota et Paulo Vieira et aux membres de mon comité de thèse : Éric Grenier, Benoit Facon, Virginie Ravigné, Géraldine Roux-Morabito et Pierre Abad.

Je tiens aussi à remercier toutes les personnes rencontrées pendant ces trois ans avec qui j'ai eu l'occasion de discuter, de boire un verre et même de découvrir l'escalade !

Je ne pourrais pas finir sans remercier ma famille... si précieuse ! Vous ne comprenez pas toujours ce que je fais mais vous me soutenez et je ne vous remercierai jamais assez pour ça ! Pouny, Mouny, Cam's, Gool... merci d'être là. Merci à Tatie Flo', Jean-Paul et Mamie pour leur accueil quand je suis arrivée... un vrai coq en pâte ! Une pensée toute particulière pour Djoule, ma couz'...merci pour ton soutien, nos discussions, ta bonne humeur et nos séances sportives... tu m'as bien fait transpirer tu sais !! Et mes amis... Bob', Lizouille, Tchoutchou, Valoche, Bastoune, Mick'... je suis de retour !! :) Et mon Chérichou...merci pour tout, merci pour nous... toujours ensemble vers l'inconnu...Idaime ma poule !



Avant-propos

Ce projet de thèse s'est intégré dans le projet collaboratif européen REPHRAME (KBBE.2010.1.4-09), qui a démarré en Février 2011 et s'est arrêté en Septembre 2014. Son objectif était de développer de nouvelles méthodes pour détecter, contrôler et éradiquer le nématode du pin (PWN). REPHRAME compte neuf partenaires de sept pays européens. Ce projet se divisait en différents modules de travail et la thèse avait en charge (i) le développement d'un ensemble de marqueurs microsatellites spécifiques au PWN et (ii) l'évaluation de la diversité génétique des populations européennes de PWN en lien avec l'invasion.

Ce manuscrit présente les analyses effectuées pour remplir les objectifs qui nous ont été assignés et un peu plus. Il suit la forme d'un manuscrit sur articles. Les articles déjà en forme ont été intégrés dans la version permise par le journal auquel ils ont été soumis (en anglais) et sont précédés par un résumé en français. Les deux derniers travaux ne sont pas rattachés à des articles et sont donc présentés en français dans leur intégralité.

Sommaire

I. LES INVASIONS BIOLOGIQUES.....	10
1. Qu'est-ce qu'une invasion biologique ?.....	10
a. Invasion, comment la définit-on ?	10
b. Invasion, processus naturel de colonisation ?.....	11
2. Pourquoi les étudier ?.....	13
a. Invasions biologiques : partie intégrante des changements globaux	13
b. Étudier les invasions pour comprendre leur succès.....	14
c. Étudier les invasions pour comprendre les systèmes naturels.....	18
3. Identifier les routes d'invasion.....	20
a. Une étape indispensable aux multiples applications	20
b. Comment les identifier ?.....	21
❖ Types de données	22
❖ Types de méthodes	22
II. ÉTUDE DE CAS : INVASION DU NÉMATODE DU PIN	25
1. Il était une fois la maladie du pin	25
2. Il était une fois l'invasion d'un nématode	29
3. Il était une fois des objectifs	32
III. TRAVAUX RÉALISÉS.....	33
1. Diversité et structure génétiques du PWN dans le monde	34
a. Mise au point des outils et première étude de l'aire native	34
b. Diversité et structure génétiques en lien avec l'invasion.....	47
2. Effet "collection"	76
3. Routes d'invasion par la méthode ABC	111
4. Explorer les raisons du succès de l'invasion	142
IV. DISCUSSION GÉNÉRALE	154
1. Invasion du nématode du pin	154
a. Principaux résultats.....	154
❖ Fonctionnement des populations.....	154
❖ Introduction et routes d'invasion	155
❖ Raisons du succès.....	157
❖ Effet "collection"	158
b. Limites et perspectives	160
2. Au-delà de ce cas d'invasion	164
a. Mesures de différenciation	164
b. Méthodes traditionnelles et méthode ABC.....	167
c. Les nématodes et la génétique des populations	168
RÉFÉRENCES BIBLIOGRAPHIQUES	170

I. LES INVASIONS BIOLOGIQUES

1. Qu'est-ce qu'une invasion biologique ?

a. Invasion, comment la définit-on ?

Une invasion biologique se définit comme la prolifération d'organismes dans de nouvelles aires géographiques, souvent distantes de leurs aires de répartition originelles (Mack et al. 2000). C'est un processus en cascade, composé de plusieurs étapes. Bien qu'un nombre différent d'étapes puisse être défini selon le niveau de précision considéré, les trois principales sont, par ordre chronologique : l'introduction, l'établissement et la prolifération (Mack et al. 2000; Sakai et al. 2001). L'introduction d'un organisme hors de son aire de répartition originelle ne définit donc pas seule une invasion biologique. L'invasion réside principalement dans la prolifération, en nombre et dans l'espace (Daehler 2001; Richardson et al. 2000). Parmi les organismes naturalisés (dont la persistance ne dépend plus d'une immigration fréquente en provenance de l'aire native), seule une partie deviendra envahissante (Mack et al. 2000). Les organismes doivent réussir chaque étape, conduisant à un faible taux de réussite des invasions biologiques, généralement illustré par la célèbre règle des 10% de Williamson & Fitter (1996). Cette règle stipule que, grosso modo, 10% des organismes réussissent à chaque étape de l'invasion, ce qui conduit à une invasion réussie sur 1000 tentatives. Loin de décrire une vérité absolue, cette règle insiste sur le fait que les invasions qui réussissent sont bien moins fréquentes que les invasions qui échouent.

On pourrait être surpris de l'absence de considération explicite des impacts potentiels des invasions biologiques dans la définition présentée ici, bien que les termes "invasion" et

"prolifération" puissent en soi traduire des impacts négatifs. Tous les biologistes ne s'accordent pas sur ce point. Davis & Thompson (2000), par exemple, proposent de qualifier d'envahissants seulement les organismes ayant un impact important sur leur nouvel environnement, indépendamment du critère d'expansion/prolifération. La littérature fourmille effectivement d'exemples documentant l'impact négatif des invasions biologiques, qui justifie bien souvent leur étude (voir partie I.2.a). Cependant, la notion d'impact est subjective, difficile à quantifier et toutes les invasions n'engendrent pas d'impacts détectables (Daehler 2001; Ricciardi & Cohen 2007; Richardson et al. 2000; Simberloff et al. 2013). Exclure cette notion de la définition permet de ne pas la restreindre et de se concentrer sur les mécanismes et les processus démographiques et génétiques à l'origine des invasions biologiques. Ces derniers sont identiques quels que soient les impacts de l'invasion, mesurables objectivement et leur compréhension est essentielle pour d'autres domaines de l'écologie (voir partie I.2.c).

b. Invasion, processus naturel de colonisation ?

La colonisation de nouvelles aires géographiques par les organismes et leur expansion sont des phénomènes naturels, qui couplés aux forces évolutives (mutation, dérive génétique et sélection naturelle), ont conduit à la biodiversité actuellement observée. Ce moteur de l'évolution est à distinguer des invasions biologiques du fait de leur fréquence, du nombre d'organismes concernés, de leur portée géographique, de leur histoire pouvant être complexe et des échelles de temps sur lesquelles elles se produisent (Ricciardi 2007). Une raison de la distinction entre processus naturels de colonisation et invasions biologiques est l'influence de l'Homme.

L'époque des grandes découvertes et les nombreuses explorations autour du monde associées ont marqué un tournant décisif dans la dispersion d'organismes à travers les grandes

barrières biogéographiques. D'impossible à rarissime, cette dernière est devenue courante voire routinière (Mooney & Cleland 2001). Le rôle prépondérant de l'homme n'est plus à démontrer. Bon nombre d'auteurs s'accordent sur le fait que le nombre des invasions biologiques a augmenté et augmente toujours en conséquence directe du commerce international et de l'amélioration des infrastructures qu'il nécessite, facilitant considérablement le mouvement des organismes et leur introduction dans de nouvelles aires (Cohen & Carlton 1998; Levine & D'Antonio 2003; Lodge & Shrader-Frechette 2003; Lounibos 2002; Mack et al. 2000; Mooney & Cleland 2001; Pyšek et al. 2010; Tatem et al. 2006; Westphal et al. 2008). La comparaison des taux de colonisation naturelle et d'invasion biologique présentée par Ricciardi (2007) appuie ce fait en estimant le taux de colonisation naturelle à 0.33 espèces/million d'années contre 35000 espèces/million d'années pour les invasions biologiques. Le cas précis de la baie et du delta de San Francisco, avec un taux d'invasion passant d'un nouvel envahisseur tous les ans de 1851 à 1960 à plus de trois par an de 1961 à 1995, illustre également cette augmentation (Cohen & Carlton 1998). De plus, au-delà des introductions fortuites, les introductions délibérées (de plantes et de vertébrés notamment) ont fortement contribué à augmenter la fréquence des invasions biologiques (Mack et al. 2000). Finalement, le réchauffement climatique mais aussi l'altération des écosystèmes, principalement l'homogénéisation environnementale, résultats des activités anthropiques, peuvent, *via* des interactions synergiques, faciliter le succès des invasions (Guillemaud et al. 2011; Ricciardi 2007; Soulé 1992; Vitousek et al. 1996).

2. Pourquoi les étudier ?

a. Invasions biologiques : partie intégrante des changements globaux

Toutes les invasions biologiques ne conduisent pas à des impacts détectables (ou négatifs selon le point de vue considéré) comme évoqué précédemment. Néanmoins, la prolifération des organismes introduits se fait bien souvent au détriment de la faune et la flore de l'aire d'introduction, justifiant la nécessité de les étudier.

Les organismes envahissants altèrent de différentes façons la trajectoire évolutive ou l'écologie des organismes natifs. On peut citer de manière non exhaustive : l'exclusion compétitive, le déplacement de niche, l'hybridation et l'introggression (entraînant des changements et des extinctions génétiques, ainsi que la perte de génotypes localement adaptés et/ou la production de nouveaux organismes hautement envahissants), la prédation et l'extinction, conduisant à la réduction de la diversité biologique (Mooney & Cleland 2001; Rhymer & Simberloff 1996; Sakai et al. 2001; Simberloff 2003). Les invasions biologiques sont également en partie responsables de l'homogénéisation biotique (McKinney & Lockwood 1999) et peuvent altérer les processus écosystémiques (Mack et al. 2000; Vitousek et al. 1996), affectant les services dont nous bénéficions (Boyd et al. 2013 pour un exemple sur les forêts). Elles peuvent enfin agir comme des vecteurs de maladies végétales, animales et humaines (Lounibos 2002; Mack et al. 2000; Soulé 1992; Vitousek et al. 1996).

Pimentel et al. (2000; 2005) ont tenté de quantifier l'impact économique de ces effets. Ainsi, chaque année, aux USA, les dommages liés aux organismes envahissants s'élèveraient à près de 120 milliards de dollars, montant probablement sous-estimé car ne prenant en compte qu'un sous-ensemble des organismes envahissants et excluant les conséquences

indirectes et non commerciales des invasions biologiques (Pimentel et al. 2000; Pimentel et al. 2005). De plus, près de 50% des espèces menacées ou en voie de disparition aux USA le sont à cause de la compétition avec et/ou de la prédation par des organismes envahissants (Wilcove et al. 1998). Comme le résumaient très bien Vitousek et al. (1996), "our mobile society is redistributing the species on the earth at a pace that challenges ecosystems, threatens human health and strains economies". De par toutes ces conséquences néfastes et leur récurrence, les invasions biologiques sont actuellement – et ce depuis plusieurs dizaines d'années – reconnues comme la deuxième menace pesant sur la biodiversité (Walker & Steffen 1997). Elles doivent être considérées au même titre que les changements d'usage des terres, l'augmentation du CO₂ et la surexploitation et font de ce fait partie intégrante des changements environnementaux globaux induits par l'Homme (Mack et al. 2000; Ricciardi 2007; Sala et al. 2000; Vitousek et al. 1996; Wilcove et al. 1998).

b. Étudier les invasions pour comprendre leur succès

Les invasions biologiques sont des phénomènes essentiellement irréversibles (Mooney & Cleland 2001; mais voir Simberloff 2009 pour des exemples d'éradication réussies). Leur nombre augmente et leurs conséquences sont nombreuses. Réussir à identifier les espèces envahissantes et à comprendre les processus conduisant à une invasion réussie constituent donc des défis énormes, tant d'un point de vue écologique, économique, sociétal que scientifique (Mack et al. 2000).

Les invasions biologiques ont un faible taux de réussite et tous les organismes introduits n'arrivent pas à envahir (Mack et al. 2000; Williamson & Fitter 1996). Il existe ainsi des variations individuelles qui traduisent des différences de potentiel envahissant (exclusion faite des déterminants extérieurs aux organismes tels que les caractéristiques de

l'aire d'introduction pouvant également influencer le succès d'une introduction). Étudier ces variations peut permettre de caractériser précisément les organismes qui ont réussi à envahir pour dégager les traits qui déterminent le succès et prédire quels organismes ont un fort potentiel envahissant. Prédire les invasions est un objectif plus que séduisant du fait de l'atout majeur que cela apporterait pour limiter les invasions futures. Cependant, comme Williamson (1999) qui suppose que les invasions peuvent être aussi prévisibles que les tremblements de terre ou encore comme Enserink (1999) qui suppose qu'établir une vraie théorie prédictive constitue "le Graal" de la biologie de l'invasion, de nombreux biologistes pensent que cet objectif séduisant n'est pas encore réaliste. Même si certains auteurs sont plus optimistes au vue des progrès réalisés (Kolar & Lodge 2001), la généralisation et le niveau de connaissances que nécessite une vraie prédiction constituent sûrement toujours des limites majeures.

Outre le désir d'identifier les organismes potentiellement envahissants, améliorer notre compréhension des invasions et tenter de les expliquer sont des objectifs tout aussi nécessaires.

Le paradoxe génétique entourant le succès d'une introduction (Allendorf & Lundquist 2003; Sakai et al. 2001) compte parmi les éléments à expliquer pour comprendre le succès des invasions biologiques. Les populations introduites sont souvent de petites tailles en raison des goulots d'étranglement et des effets fondateurs qu'implique l'introduction. Elles sont donc généralement moins diversifiées que les populations natives dont elles sont issues. Cette réduction de diversité génétique subie par les populations introduites peut avoir des conséquences importantes sur (i) leur persistance et leur croissance *via* le phénomène de dépression de consanguinité et sur (ii) leur capacité à évoluer dans leur nouvel environnement. Expliquer ce paradoxe génétique revient ainsi à comprendre comment une population, qui a certainement subi un goulot d'étranglement lors de son introduction, peut

survivre dans un nouvel environnement, s'y adapter et l'envahir, en dépit de la perte de diversité génétique qu'elle a subie et des conséquences négatives que cela entraîne.

Plusieurs hypothèses ont été avancées pour tenter de l'expliquer et certaines d'entre elles ont déjà été testées et sont bien documentées. C'est notamment le cas des introductions multiples dans la même zone géographique (Frankham 2005). Elles permettent de compenser la perte de diversité génétique liée à chaque introduction, voire d'augmenter la diversité de la population envahissante. Les études menées sur des exemples d'invasion ont ainsi montré que de nombreux cas sont issus d'introductions multiples (Bossdorf et al. 2005 pour une revue chez les plantes; Facon et al. 2008; Kelager et al. 2013; Kolbe et al. 2004; Miller et al. 2005; Roman & Darling 2007 pour une revue chez les organismes aquatiques). Il semblerait même que leur forte fréquence remette en cause l'existence du paradoxe, qui pour certains ne tiendrait plus (Roman & Darling 2007). Ensuite, l'hybridation, qui peut-être fréquente dans le cas d'introductions multiples, est un autre mécanisme proposé favorisant les invasions (Ellstrand & Schierenbeck 2000; mais voir Rhymer & Simberloff 1996). Elle peut notamment provoquer le phénomène de vigueur hybride (ou heterosis, Shull 1948) et créer de nouvelles combinaisons alléliques, de nouveaux génotypes, localement plus adaptés, voire plus envahissants, que ceux des parents (Ellstrand & Schierenbeck 2000). Bien que plusieurs études vérifient cette hypothèse (Bleeker 2003; Drake 2006; Facon et al. 2005; Keller & Taylor 2010), certains chercheurs pensent que son rôle pour faciliter les invasions pourrait être surestimé (Chapple et al. 2013 pour un exemple chez un lézard; Whitney et al. 2009 pour une revue chez les plantes).

Néanmoins, de nombreuses invasions réussissent malgré une perte de diversité génétique avérée dans les populations envahissantes par rapport à leurs populations sources natives, illustrant bien le paradoxe génétique des invasions (Dlugosch & Parker 2008 pour une revue sur les effets fondateurs lors des invasions; Dlugosch & Parker 2008; Fontaine et al.

2013; Golani et al. 2007; Papura et al. 2012; Perdereau et al. 2013 pour des exemples). D'autres explications, au-delà des mécanismes compensant la perte, ont donc été proposées. Il a été proposé qu'un goulot d'étranglement modéré pouvait induire la purge des allèles délétères, réduisant ainsi la dépression de consanguinité que peut subir les petites populations après leur introduction, et facilitant l'expansion de la population envahissante (Facon et al. 2011). Il est aussi possible que la perte de diversité génétique lors de l'introduction ne soit pas en soi un problème dans la mesure où les populations envahissantes n'ont pas nécessairement à s'adapter après leur arrivée. Cela peut être (i) le cas lors d'une invasion via le phénomène "tête de pont", lorsqu'une population envahissante est elle-même à l'origine d'autres foyers envahissants (Lombaert et al. 2010), (ii) illustré par l'hypothèse de l'"AIAI" (pour anthropogenically induced adaptation to invade), lorsque des phénomènes d'adaptation ont eu lieu dans l'aire native et ont prédisposé les organismes à réussir leur invasion dans l'aire d'introduction (Hufbauer et al. 2012 pour la conceptualisation de ce phénomène; Foucaud et al. 2009; Bossdorf et al. 2008 pour des exemples) ou (iii) expliqué par une aire d'introduction similaire à l'aire native, qui présente les mêmes conditions climatiques (Tatem & Hay 2007), sans ennemis naturels (prédateurs/parasites/pathogènes, Mitchell & Power 2003; Prenter et al. 2004; Torchin et al. 2003) et/ou avec des hôtes sensibles. Enfin, un cas célèbre d'invasion (celle de la fourmi argentine, Tsutsui et al. 2000) démontre même comment l'absence de diversité peut faciliter, voire être indispensable à la réussite d'une invasion biologique.

Les connaissances sur les invasions s'accumulent et plusieurs cas d'invasion ont pu être expliqués. Cependant, certaines hypothèses manquent de preuves et comme pour la prédiction, l'explication souffre d'une généralisation difficile. Des études de cas supplémentaires, qui augmentent la diversité des organismes impliqués, des histoires d'invasion et des écosystèmes concernés, sont donc nécessaires pour éprouver les hypothèses

avancées et continuer à améliorer nos connaissances des invasions biologiques pour mieux les expliquer.

c. Étudier les invasions pour comprendre les systèmes naturels

Les exemples d'invasion constituent des éléments utiles à la compréhension des raisons du succès des invasions. Il s'agit également d'une expérimentation *in natura* (Grinnell 1919) qui peut être utilisée pour étudier et comprendre les forces à l'œuvre dans les systèmes naturels (Sax et al. 2007). Les exemples d'invasions offrent ainsi de bonnes opportunités pour les recherches dans différents domaines de l'écologie. Levin (2003) souligne par exemple l'utilité des invasions pour comprendre la spéciation, du fait de la ressemblance de ces processus qui nécessitent tous deux une étape d'introduction et une étape d'établissement. Également, Westley (2011) reconnaît les invasions comme des modèles opportunistes pour tenter de déterminer le taux et la forme (rapide ou lente) des changements phénotypiques dans les populations naturelles. Sax et al. (2007) présentent aussi dans leur revue, l'atout que représentent les espèces envahissantes pour identifier les mécanismes qui entraînent l'adaptation au nouvel environnement. Dans ce cas précis, l'étude des invasions biologiques a par exemple mis en évidence que les réductions de diversité génétique ne semblent pas limiter systématiquement les capacités d'adaptation des populations introduites dans le nouvel environnement (Dlugosch & Parker 2008; Dlugosch & Parker 2008; Wares et al. 2005). Enfin Puth & Post (2005) insistent aussi sur l'intérêt d'étudier les invasions, et particulièrement leurs premières étapes, en raison de leur importance pour comprendre comment les communautés s'assemblent et se structurent et comment elles répondent et récupèrent après une perturbation.

La dynamique des invasions biologiques étant en partie comparable à celle recherchée dans le cadre de la lutte biologique ou de la réintroduction d'espèces menacées, l'étudier peut aussi nous aider à améliorer leur succès. En effet, certains chercheurs soulignent l'utilité de voir les organismes envahissants comme des organismes modèles. Leur étude peut aider à développer des stratégies de gestion et conservation *via* la compréhension de l'extinction des individus ou espèces, le fonctionnement des écosystèmes et les réponses des individus ou espèces aux changements climatiques (Sax et al. 2007). De même, de nombreuses questions non résolues concernant l'application de la génétique à la conservation, comme les effets de la consanguinité des petites populations ou l'importance de l'adaptation locale, peuvent être traitées grâce aux invasions biologiques (Allendorf & Lundquist 2003).

Enfin, étudier l'expansion des populations envahissantes est également important pour améliorer nos connaissances théoriques sur les processus de colonisation/dispersion. Les invasions et plus particulièrement l'expansion post-introduction soulèvent en effet des questionnements concernant la structuration génétique spatiale des populations invasives dans leur nouvel environnement. Etudier l'expansion des populations envahissantes peut ainsi permettre de caractériser la mise en place récente d'un patron spatial de structure génétique et d'en déduire le modèle d'expansion : diffusion continue, migration longue distance ou diffusion stratifiée (décrite initialement dans Shigesada et al. 1995).

Comme concluaient Williamson & Fitter (1996), la question est de savoir si l'étude de la dynamique des populations contribue plus à la compréhension des invasions biologiques que l'étude des invasions biologiques ne contribue à la compréhension de la dynamique des populations.

3. Identifier les routes d'invasion

La compréhension des invasions biologiques passe par la compréhension de l'histoire du processus d'invasion, incluant notamment l'identification des routes d'invasion. Les routes d'invasion se définissent comme les chemins suivis par les organismes (individus, graines,...) entre leurs populations sources et les populations envahissantes qu'ils ont formées dans l'aire d'introduction. Identifier les routes d'invasion est un pré-requis important comme nous allons le voir dans le paragraphe suivant.

a. Une étape indispensable aux multiples applications

La reconstruction des routes d'invasion est utile pour des raisons appliquées et théoriques (comme explicitées par Estoup & Guillemaud 2010). D'un point de vue appliqué, l'identification des routes d'invasion permet de prévenir les invasions car elle facilite la conception et la mise en place de mesures pour limiter l'introduction, l'établissement ou l'expansion. Tout d'abord, elle permet potentiellement de prévenir les nouvelles introductions en améliorant la vigilance envers les populations source identifiées *via* une surveillance accrue et des mesures de quarantaine (Mack et al. 2000). Ensuite, couplée à des informations sur le niveau de connectivité des populations envahissantes (obtenues par des analyses génétiques par exemple), l'identification des routes d'invasion permet également de renforcer ou de mieux adapter les moyens de lutte (voir Rollins et al. 2009 pour l'exemple de l'étourneau sansonnet). En effet, connaître précisément l'histoire des invasions permet d'obtenir des informations directes sur l'origine des populations envahissantes et sur l'éventualité d'introductions multiples (par ex., Ciosi et al. 2008; Miller et al. 2005). Ces informations sont utiles pour définir des mesures de contrôle efficaces. Par exemple, elles

aident à identifier les ennemis naturels les plus efficaces pour la lutte biologique, en permettant de se concentrer sur la source des populations envahissantes (Strong & Pemberton 2000; van Wilgen et al. 2013). Enfin, l'identification des routes d'invasion permet aussi de définir les caractéristiques écologiques des organismes envahissants – leur biologie et leur niche écologique – en les étudiant dans leur aire native. Cela constitue un atout pour la mise en place de mesures de contrôle efficaces et pour prédire et anticiper la distribution potentielle des organismes envahissants dans l'aire d'introduction (Kolar & Lodge 2001; Mack et al. 2000; Tsutsui et al. 2000).

D'un point de vue théorique, une connaissance précise des routes d'invasion permet d'établir et de tester les hypothèses écologiques ou évolutives avancées pour expliquer le succès des invasions (Estoup & Guillemaud 2010). En effet, connaître la population native source d'une population envahissante rend possible la réalisation d'expériences de laboratoire pour comparer les deux types de populations. Ces comparaisons entre population native et population envahissante correspondante sont importantes et utiles pour tester, identifier et mesurer les facteurs qui peuvent avoir évolué pendant l'invasion, expliquant potentiellement son succès (Facon et al. 2006; Keller & Taylor 2008).

b. Comment les identifier ?

Différentes types de données et méthodes sont traditionnellement utilisées pour identifier les routes d'invasions. Estoup & Guillemaud (2010) en fournissent une présentation, reprise ici après modification.

❖ Types de données

Deux grands types de données, informatives pour retracer les routes d'invasion, existent : les données dites "directes" et les données dites "indirectes". Les données dites "directes" correspondent à des informations de type présence/absence des organismes à un endroit donné à différentes dates. Ces informations sont obtenues grâce aux contrôles réguliers réalisés dans les aéroports ou ports de commerce par les agences nationales environnementales, agricoles ou s'occupant des organismes de quarantaine, ainsi qu'aux observations faites par des naturalistes et/ou des amateurs dans le cadre des sciences participatives par exemple. Ces données seules sont limitées pour retracer précisément les routes d'invasions. En effet, les relations telles que "la population A est la source de la population B" ne sont pas simplement déductibles de ces données. De plus, au vu du faible taux de réussite des invasions, rien n'assure que les individus interceptés auraient envahi avec succès (mais voir Tatem et al. 2006).

Les données dites "indirectes" sont issues de marqueurs génétiques, tels que les marqueurs microsatellites par exemple. Elles correspondent à des génotypes individuels dont l'analyse fournit des informations sur les relations d'apparentement des individus, sur la structure des populations et sur leur histoire évolutive. En particulier, les données génotypiques donnent des informations sur les divergences entre populations, sur l'intensité de la dérive, et sur l'existence de migration entre populations (Hamilton 2009). Elles peuvent donc permettre d'inférer l'histoire des populations.

❖ Types de méthodes

On peut distinguer d'une part les méthodes descriptives ou qualitatives, regroupant les méthodes d'analyses classiques de la génétique des populations, et d'autre part les méthodes quantitatives ou basées sur des modèles, parmi lesquelles la méthode ABC (pour

"Approximate Bayesian Computation"). Cette dernière est une méthode statistique récente (Beaumont et al. 2002) qui, appliquée aux invasions biologiques, constitue une avancée majeure pour identifier les routes d'invasion (Miller et al. 2005; Pascual et al. 2007).

Les méthodes descriptives utilisent exclusivement les données dites "indirectes" et reposent sur l'analyse des patrons génétiques observés au sein et entre les populations. Elles incluent entre autres la construction d'arbres à partir de distances génétiques (Saitou & Nei 1987 pour un exemple de méthode), le calcul de vraisemblances d'assignation (Paetkau et al. 2004; Rannala & Mountain 1997) et les mesures de différenciation entre populations (Jost 2008; Weir & Cockerham 1984). Plus récemment, des méthodes Bayésiennes de groupement d'individus (comme celle implémentée dans le logiciel Structure, Pritchard et al. 2000) ont également été utilisées pour tenter d'identifier les routes d'invasion des populations envahissantes et notamment pour déterminer leur source (Ascunce et al. 2011; Perdereau et al. 2013; Rollins et al. 2009). L'avantage majeur de ces méthodes réside dans leur simplicité conceptuelle, et leur facilité et rapidité de réalisation. Elles ont également montré leur efficacité pour retracer les routes dans plusieurs cas d'invasions (Ciosi et al. 2008; Facon et al. 2003; Papura et al. 2012; Perdereau et al. 2013; Wan et al. 2012). Elles présentent néanmoins des limites : (i) les données génétiques sont réduites à une ou quelques statistiques résumant la variation génétique entre les populations, (ii) elles ne prennent pas en compte la stochasticité de l'histoire démographique et génétique des invasions ni la stochasticité d'échantillonnage et (iii) elles ne permettent pas de comparer statistiquement, *via* le calcul de probabilités, différents scénarii d'invasions.

Les méthodes quantitatives, fondées sur des modèles, permettent de comparer quantitativement des scénarii. C'est le cas de l'ABC qui pallie les limites sus-citées et permet de combiner les données dites "directes" et "indirectes". Elle est de ce fait utilisée dans de nombreuses études sur les routes d'invasion (Ascunce et al. 2011; Barres et al. 2012; Boissin

et al. 2012; Boubou et al. 2012; Lombaert et al. 2010; Miller et al. 2005; Pascual et al. 2007; Rius et al. 2012). L'ABC a originellement été développée pour estimer des paramètres de génétique des populations (Beaumont et al. 2002) mais elle a séduit dans divers domaines de la biologie (par exemple en phylogéographie, conservation et épidémiologie, Beaumont et al. 2010; Lopes & Boessenkool 2010; Lopes & Beaumont 2010). L'ABC est une méthode d'inférence basée sur des modèles, ancrée dans les statistiques Bayésiennes. Elle est utilisée lorsque les modèles testés sont trop complexes pour que la fonction de vraisemblance soit calculée. En effet, elle permet, grâce à des simulations intensives, de contourner le calcul de la fonction de vraisemblance (Bertorelle et al. 2010; Csillery et al. 2010; Sunnaker et al. 2013). Dans le domaine des invasions biologiques et particulièrement dans son utilisation pour retracer les routes d'invasion, l'ABC permet (i) de prendre en compte les connaissances biologiques et historiques (c'est ici que ce fait l'utilisation des données "directes" de présence-absence) du modèle étudié lors de la définition des distributions *a priori* des paramètres (démographiques et mutationnels) et dans le choix des scénarii testés, (ii) de tester des scénarii d'invasion complexes (par ex., introductions multiples, hybridation), (iii) de traiter les échantillonnages incomplets *via* l'intégration de populations "fantômes", (iv) de ne pas (trop) réduire les données génétiques en utilisant plusieurs statistiques pour résumer la variation génétique entre les populations, (v) de comparer statistiquement, *via* le calcul de probabilités *a posteriori*, différents scénarii d'invasions et (vi) d'évaluer la confiance que l'on peut avoir dans le choix du scénario, grâce aux calculs d'erreurs de type I et de type II, aux intervalles de confiance des probabilités *a posteriori* et à la vérification de l'adéquation du modèle sélectionné aux données.

Le fonctionnement de l'ABC est repris en détails dans la partie III.3.

II. ÉTUDE DE CAS : INVASION DU NÉMATODE DU PIN

Durant cette thèse, nous avons étudié le cas de l'invasion du nématode du pin, *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae). Nous présentons dans cette partie le modèle étudié et la maladie qu'il provoque. Le nématode du pin a un cycle de vie relativement complexe d'où la nécessité de le décrire.

1. Il était une fois la maladie du pin

Les nématodes sont des vers ronds non segmentés, généralement inférieurs à 1 mm de long. Ils sont ubiquistes et sont retrouvés dans les écosystèmes terrestres aussi bien que marins. Ils présentent une diversité de stratégies de vie, étant de formes libres ou parasites de plantes, d'animaux et de l'Homme. Parmi les 26 000 espèces décrites, environ 15 % sont des nématodes parasites de plantes (Hugot et al. 2001). Le nématode du pin (en anglais, pinewood nematode, PWN), *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae) fait partie de ces 15 %. Il est responsable de la maladie du dépérissement du pin (en anglais pine wilt disease, PWD, Mamiya 1972; Mamiya 1976; Mamiya 1983). La PWD est le résultat d'un processus biologique complexe (schématisé par la figure 1, ci-dessous) qui combine au moins trois acteurs : le nématode, l'insecte vecteur et l'arbre sensible, présentés sur la photo 1 (Evans et al. 1996; Futai 2013; Jones et al. 2008).

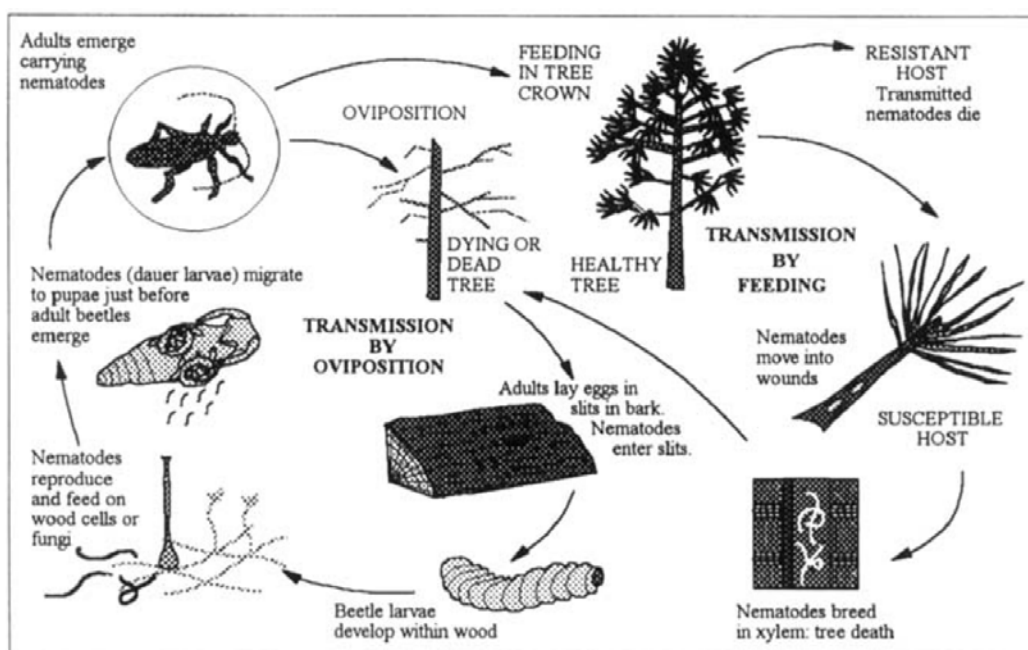


Figure 1 : Relation entre le PWN, l'insecte vecteur et l'arbre (Evans et al. 1996)

Les insectes vecteurs du PWN sont quasi exclusivement les longicornes du genre *Monochamus* (Coleoptera : Cerambycidae, Akbulut & Stamps 2012; Linit 1988). Les principales espèces vectrices sont : *Monochamus carolinensis* Olivier en Amérique du Nord (Linit 1988), *Monochamus alternatus* Hope en Asie (Kobayashi et al. 1984; Mamiya & Enda 1972) et *Monochamus galloprovincialis* Olivier au Portugal (Naves et al. 2007; Sousa et al. 2001; Sousa et al. 2002), potentiellement en Espagne (Robertson et al. 2011) et en Europe de manière générale (Akbulut & Stamps 2012).



Photo 1 : Les trois acteurs principaux impliqués dans la PWD. De gauche à droite, le PWN (inférieur à 1mm), l'insecte vecteur (de 15 à 30 mm) et l'arbre sensible.

En ce qui concerne l'arbre, plusieurs espèces de conifères sont sensibles au PWN mais le genre *Pinus* est le plus touché (Evans et al. 1996), *Monochamus* sp. se nourrissant principalement de ce genre de conifères (Akbulut & Stamps 2012). Il existe différents niveaux

de sensibilité et les différentes espèces de pin sont généralement classées en 3 catégories : sensibles, intermédiaires ou résistantes. Les plus sensibles sont (liste non exhaustive) : *P. thunbergii*, *P. luchuensis*, *P. densiflora* et *P. massoniana* en Asie ainsi que *P. pinaster*, *P. sylvestris* et *P. nigra* en Europe (Evans et al. 1996; Fielding & Evans 1996).

Le PWN est mycophage et phytophage. Il peut donc se nourrir d'arbres vivants. Il existe ainsi deux modes de transmission pour cette espèce (Figure 1, ci-dessus) : transmission par oviposition sur arbres morts ou mourants par les femelles et transmission par alimentation

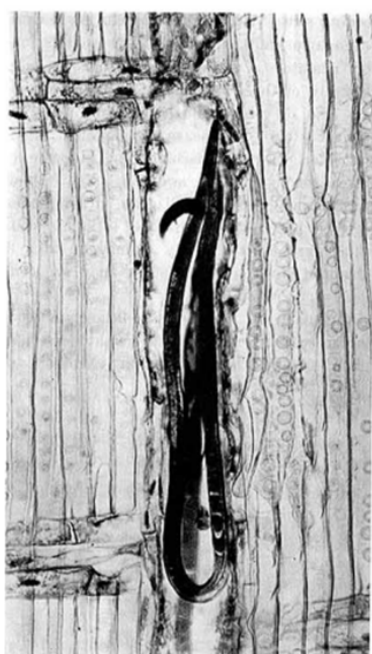


Photo 2 : PWN dans un canal de résine. Les cellules épithéliales sont détruites (tirée de Mamiya 1976)

de maturation par les insectes émergents sur arbres sains et vivants (Fielding & Evans 1996; Jones et al. 2008; Mamiya 1983). C'est ce second mode de transmission qui conduit à la PWD, le PWN étant alors responsable de la mort de l'arbre (Li et al. 2007; Naves et al. 2007). Les nématodes profitent des blessures causées par les longicornes pendant leur alimentation de maturation pour pénétrer dans l'arbre et se répandre dans les tissus vasculaires et canaux de résine (Photo 2). Ils se nourrissent des cellules épithéliales qui tapissent les canaux de résine, des champignons qui envahissent l'arbre lorsque celui-ci flétrit, et se reproduisent à l'intérieur de l'arbre, bloquant ainsi la conductance de l'eau et provoquant la mort

de l'arbre (Bergdahl 1988; Mamiya 1983). Le cycle s'auto-entretient donc, l'arbre infesté devenant propice à l'oviposition des femelles longicornes (Figure 1).

Plus précisément, le PWN se reproduit sexuellement (Futai 2013) et se développe en quatre stades larvaires (J1 à J4) suivi de la vie d'adulte sexuée. Il présente deux phases alternatives dans son cycle de vie : une phase reproductive/multiplicative et une phase dispersante/résistante (dont les larves sont appelées des larves "dauer", signifiant larves de

résistance), présentées dans la figure 2 ci-contre. Les deux premiers stades larvaires sont communs aux 2 phases et se déroulent dans l'œuf. Pendant la première phase, le PWN se reproduit rapidement et envahit l'arbre lorsque les conditions sont favorables. Pendant la seconde phase, il est

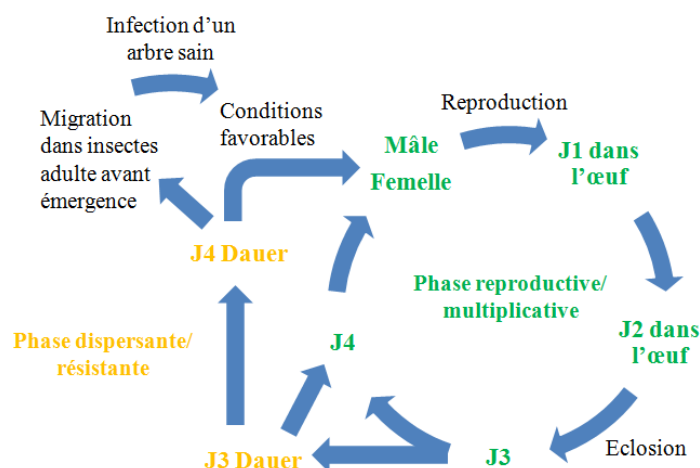


Figure 2 : Cycle biologique du PWN

transporté par le vecteur mais peut aussi résister à des conditions moins clémentes (Evans et al. 1996; Warren & Linit 1993). La durée du cycle de vie du PWN est fortement influencée par la température, que ce soit en laboratoire (en culture sur *Botrytis cinerea*, Mamiya 1975) ou dans la nature (dans l'arbre, Mamiya & Enda 1972). Ainsi, plus il fait chaud (jusqu'à une température maximale seuil de 33°C), plus le cycle est rapide : 4-5 jours à 25°C versus 12 jours à 15°C, sur *Botrytis cinerea* (Mamiya 1975). Une conséquence de cette influence environnementale a été mise en évidence par Kiyohara & Tokushige en 1971 (dans Mamiya 1983). En effet, ces auteurs ont montré qu'une inoculation en été produisait une apparition plus rapide des symptômes de la maladie qu'une inoculation de Février à Mai. Ils ont également souligné qu'une inoculation en Septembre/Octobre ne provoquait aucun symptôme. Le développement chronologique du PWN et donc de la PWD coïncide parfaitement avec le cycle de vie de l'insecte vecteur (Mamiya & Enda 1972), comme présenté ci-dessous sur la figure 3.

Au-delà du blocage mécanique causé par la reproduction du PWN, l'hypothèse d'association avec des bactéries a été avancée pour expliquer la pathogénicité du PWN. Cette hypothèse a été vérifiée dans plusieurs études. Néanmoins, d'autres études ont également

montré que le PWN était capable de provoquer la PWD en condition aseptique (controverse présentée dans Vicente et al. 2011)

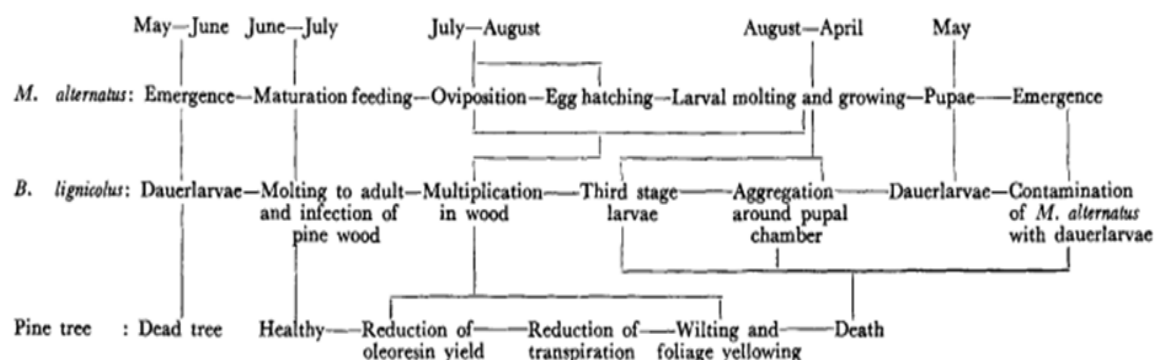


Figure 3 : Relations chronologiques et biologiques entre *Bursaphelenchus xylophilus* (initialement appelé *Bursaphelenchus lignicolus*), l'insecte vecteur japonais et l'hôte (Mamiya 1972).

2. Il était une fois l'invasion d'un nématode

L'Amérique du Nord est généralement reconnue comme l'aire native du PWN et il est largement répandu aux USA et au Canada (Dropkin et al. 1981; Dwinell & Nickle 1989; Kiritani & Morimoto 2004; Wingfield et al. 1982). Dans son aire originelle, il n'est pas pathogène sur les espèces locales de conifères mais tue les espèces exotiques, notamment *Pinus sylvestris* (Bergdahl 1988; Dwinell 1985; Dwinell 1997; Evans et al. 1996).

Les nématologistes pensent que le PWN a été introduit au Japon en 1905, près de Nagasaki sur l'île de Kyushu (Mamiya 1988). Ils ont déterminé cette année d'introduction car elle coïncide avec les premières observations de dépérissement massif de pins, le PWN n'ayant été identifié qu'en 1971 comme responsable de la PWD. En 1981, la perte annuelle de pins japonais s'élevait à 2 millions de m³, correspondant à environ 10 millions d'arbres (Mamiya 1983; Mamiya 1988). La zone infestée par la maladie représentait 28% des 2,1 millions d'hectares de forêts de pins existant à l'échelle nationale japonaise en 2000 (Mamiya

2004). D'autres pays asiatiques ont signalé l'arrivée du PWN sur leur territoire à partir des années 1980 : la Chine en 1982, Taïwan avant 1985 et la Corée du Sud en 1988 (Futai 2013). Là également, des dommages importants sur les pins natifs ont été rapportés. Dans toute la Chine, la PWD a infesté un million d'hectares de forêts et 44 millions de pins entre 1982 et 2007 (Yu et al. 2011). En 1999, le PWN a été identifié pour la première fois en Europe sur des pins maritimes, dans la région de Setubal au Portugal (Mota et al. 1999). De nouvelles émergences ont été signalées dans le centre du Portugal en 2008 malgré les actions gouvernementales d'éradication mises en place (Rodrigues 2008), sur l'île de Madère en 2009, près de Funchal (Fonseca et al. 2012), ainsi qu'en Espagne, en 2008 et 2010, respectivement dans les régions d'Estrémadure et de Galice (Abelleira et al. 2011; Robertson et al. 2011).

Le PWN est considéré comme une menace sérieuse pour les forêts de pins mondiales en raison des conséquences écologiques et économiques qu'engendre son introduction (Mamiya 1988; Soliman et al. 2012; Suzuki 2002; Vicente et al. 2011). À noter cependant que dans le cas de la Chine, Yu et al. (Yu et al. 2011) proposent de tirer avantage de ce désastre. Ils suggèrent en effet que le PWN accélère la succession des plantations/forêts de pins par des forêts de feuillus à feuilles persistantes, qui fournissent plus de services écologiques tels que le stockage de carbone et la résilience écologique. Le PWN a été désigné organisme de quarantaine par l'Organisation Européenne et méditerranéenne pour la Protection des Plantes depuis 1985 (OEPP/EPPO, 2014). Des mesures phytosanitaires strictes ont été imposées sur le commerce de bois de conifères entre états membres de l'UE (depuis son arrivée en Europe, Commission Européenne 2006 ; Commission Européenne 2009) mais également en provenance des pays infestés hors UE et de l'aire native. Ces mesures obligent notamment les pays à traiter et à contrôler le bois importé ou exporté, ajoutant des coûts significatifs pour les pays infestés et non infestés. De plus, depuis le début des années 80's, des restrictions

européennes d'importation de bois nord-américain engendrent également des coûts importants pour l'aire native du PWN (Bergdahl 1988; Dwinell & Nickle 1989; Liebhold et al. 1995).

Les moyens officiels de contrôle sont assez lourds et radicaux et représentent également des coûts importants. Ils passent par l'abattage et la destruction rapide des arbres infestés et la création de zones "tampon" (de 500 m à 3 km de rayon autour des arbres infestés), dans lesquelles la coupe franche d'arbres sensibles et une surveillance accrue est systématique (OEPP/EPPO 2012; Rodrigues 2008 pour l'exemple portugais). Détecter rapidement de nouveaux arbres infestés n'est pas une tâche aisée car il n'existe pas de moyens pour identifier visuellement les arbres mourants à cause du PWN, de fait de la non-spécificité des symptômes. Des méthodes de détection rapides et précises ont donc été développées afin d'améliorer et de faciliter la surveillance. Elles permettent par exemple d'éviter l'étape d'extraction des nématodes en vérifiant directement la présence du PWN dans un échantillon de bois ou dans l'insecte vecteur (Cardoso et al. 2012). Au-delà des méthodes officielles, d'autres moyens de lutte et de contrôle ont été et sont toujours étudiés et testés : la lutte biologique contre le PWN à l'aide de champignons (Wang et al. 2011) ou de produits naturels dérivés des plantes (huiles essentielles et eaux de décoction par exemple, Faria et al. 2013), la lutte contre l'insecte vecteur (captures et insecticides, Evans et al. 1996) et les programmes de sélection d'arbres résistants (Ribeiro et al. 2012).

Approfondir nos connaissances sur ce cas d'invasion et sur le processus biologique complexe qui en est à l'origine permettra d'améliorer et/ou de mieux adapter les méthodes actuellement en vigueur.

3. Il était une fois des objectifs

Globalement, l'objectif de ce projet de thèse est d'affiner nos connaissances sur l'invasion du PWN pour permettre de mieux anticiper les risques d'expansion dans les forêts européennes. Précisément, cet objectif se décline en plusieurs axes de recherche, et questions correspondantes, listés ci-dessous :

- **Étude de l'aire native :** comment fonctionnent les populations de PWN ? À quelles échelles se structurent-elles ? Quel est le mode de transmission locale majoritaire ?
- **Identification des routes d'invasion :** quelle(s) est (sont) la (les) population(s) "source" des populations retrouvées en Asie et en Europe ? Y a-t-il eu un ou plusieurs évènements d'introduction en Europe ?
- **Étude de l'évolution post-introduction :** comment expliquer le succès de l'invasion du PWN ? Les populations envahissantes présentent-elles des caractéristiques phénotypiques différentes des populations natives, pouvant expliquer l'invasion ?

III. TRAVAUX RÉALISÉS

Pour tenter de comprendre le fonctionnement des populations de PWN et clarifier ses routes d'invasion, nous avons étudié la diversité et la structure génétiques des populations de PWN à travers le monde. Nous avons utilisé des méthodes traditionnelles de génétique des populations et la méthode ABC (décrites précédemment dans la partie I.3.b) pour (i) caractériser génétiquement les populations natives et envahissantes de PWN et (ii) établir les relations existantes entre les différentes populations pour déterminer *in fine* le scénario d'invasion le plus probable. Deux études se rapportent à ses analyses. Elles sont présentées dans les parties III.1.b et III.3. Ces études ont nécessité au préalable la mise au point de marqueurs microsatellites. Cette mise au point est présentée dans la partie III.1.a.

Également, lors du développement des marqueurs microsatellites, nous avons constaté qu'il existait des différences génétiques entre souches de collection et échantillons issus de populations naturelles (voir partie III.1.a). Nous avons choisi d'approfondir ce point et l'étude qui s'y rapporte est présentée dans la partie III.2.

Finalement, nous avons exploré les raisons du succès de l'invasion du PWN en comparant la croissance sur champignon (soit au laboratoire) de différents échantillons de populations naturelles natives et envahissantes. Cette exploration préliminaire est décrite dans la partie III.4.

1. Diversité et structure génétiques du PWN

dans le monde

Cette partie présente l'analyse de différents échantillons de populations naturelles natives et envahissantes de PWN qui a été réalisée *via* des méthodes traditionnelles de génétique des populations. Elle présente également la mise au point des marqueurs microsatellites, indispensables à la réalisation de cette analyse.

La mise au point des marqueurs et l'analyse de différents échantillons mondiaux de PWN sont décrites dans les articles suivants :

- Mallez S, Castagnone C, Espada M, Vieira P, Eisenback JD, Mota M, Guillemaud T, Castagnone-Sereno P. (2013). First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. PLoS ONE 8(3): e59165. doi:10.1371/journal.pone.0059165.
- Mallez S, Castagnone C, Espada M, Vieira P, Eisenback J D, Harrell M, Mota M, Aikawa T, Akiba M, Kosaka H, Castagnone-Sereno P, Guillemaud T (2014). Worldwide invasion routes of the pinewood nematode: what can we infer from population genetics analyses? Biological Invasions. doi: 10.1007/s10530-014-0788-9 (online first).

Ces articles sont résumés ci-dessous et la version publiée ou acceptée de chaque article est présentée à la suite de son résumé.

a. Mise au point des outils et première étude de l'aire native

❖ Contexte et objectifs

Comme évoqué précédemment, le PWN est une menace sérieuse pour les forêts européennes de conifères au vu des conséquences économiques et écologiques que son introduction engendre. Plusieurs études se sont déjà intéressées à l'histoire de l'invasion du

PWN. Elles ont par exemple tenté de déterminer l'origine des populations asiatiques (Cheng et al. 2008; Jung et al. 2010; Jung et al. 2010; Zhou et al. 2007) et européennes (Valadas et al. 2012; Vieira et al. 2007). Ces études souffraient cependant de plusieurs limitations, telles que l'extraction d'ADN à partir de groupes d'individus ou l'utilisation d'un nombre limité de marqueurs peu résolutifs. Il était donc important de mettre au point des marqueurs moléculaires ayant fait leurs preuves en génétique des populations (Luikart & England 1999; Selkoe & Toonen 2006) et dans l'inférence des routes d'invasions (Estoup & Guillemaud 2010), comme les microsatellites, pour clarifier précisément les routes d'invasion du PWN.

Le principal objectif de cette étude était donc de développer un ensemble de marqueurs microsatellites spécifiques au PWN et utilisables en routine. Des échantillons de populations naturelles natives ayant été utilisés pour tester le polymorphisme des marqueurs potentiels, cet article apporte également un premier aperçu de la diversité génétique et du niveau de structuration génétique existants dans l'aire native.

❖ Matériel et méthodes

Quatre-vingt quatorze paires d'amorces, devant permettre d'amplifier des séquences de plus de 80 paires de bases et correspondant à des microsatellites parfaits ou imparfaits d'au moins cinq répétitions, ont été identifiées *in silico* à partir d'une banque génomique enrichie en motifs microsatellites spécifique au PWN (Malausà et al. 2011). Nous avons testé l'amplification de chacun de ces 94 marqueurs par PCR sur 18 individus de 6 zones géographiques différentes (souches de collection, voir Tableau 1 dans l'article). Les marqueurs qui ont présenté une bonne amplification à la taille attendue ont été conservés et leur polymorphisme a été évalué grâce à 4 échantillons de populations naturelles natives et une souche de collection de l'aire native (voir Tableau 1 dans l'article). Des marqueurs

microsatellites issus de la littérature ont également été testés pour leur amplification et leur polymorphisme.

Nous avons évalué le niveau de diversité génétique des 4 échantillons de terrain et de la souche de collection en calculant les hétérozygoties attendue et observée, ainsi que le nombre d'allèles, par marqueur et globalement. Nous avons également testé si les différents marqueurs étaient à l'équilibre de Hardy-Weinberg et de liaison. Les niveaux de diversité génétique des échantillons de terrain et de la souche de collection ont enfin été comparés. En ne conservant que les échantillons de terrain, nous avons étudié la structure génétique dans l'aire native par différentes méthodes : (i) test de différenciation génétique entre les échantillons, (ii) analyse de variance moléculaire et (iii) analyse de groupement des individus par méthodes Bayésienne (Pritchard et al. 2000) et multivariée (Jombart et al. 2010).

❖ Résultats

Parmi les 94 paires d'amorces testées, 25 ont donné une bonne amplification à la taille attendue. Quinze parmi les 25 marqueurs conservés et deux marqueurs issus de la littérature ont pu être génotypés sans ambiguïté et sont apparus polymorphes. Ces 17 marqueurs ont pu être amplifiés en 3 réactions PCR multiplexées.

L'analyse des échantillons de populations naturelles natives et de la souche de collection de l'aire native a mis en évidence une diversité génétique faible à modérée. La souche de collection est apparue moins diversifiée génétiquement que les échantillons de terrain, bien que toutes les comparaisons ne soient pas apparues significatives. Finalement, les échantillons de terrain sont apparus significativement différenciés et les analyses de groupement des individus (méthodes Bayésienne et multivariée) ont suggéré l'existence de 3 groupes génétiques parmi les 4 échantillons de terrain.

❖ Discussion

Grâce à cette étude, nous avons développé 15 marqueurs microsatellites utilisables facilement en routine *via* 3 réactions PCR multiplexées. Il s'est avéré que l'un des marqueurs issus de la littérature (parmi les deux sélectionnés) était probablement le même que l'un des nôtres. Il sera donc exclu et un total de 16 marqueurs sera conservé pour les études ultérieures. Ces marqueurs ont mis en évidence une diversité génétique faible à modérée mais suffisante pour réaliser des analyses de génétique des populations. Ce nouvel ensemble de marqueurs constitue ainsi un outil utile et approprié grâce à sa résolution au niveau individuel.

Les différences génétiques entre souches de collection et échantillons de terrain ont été analysées plus précisément et les résultats sont présentés dans la partie III.2.

Nous avons enfin mis en évidence une forte structure génétique spatiale dans l'aire native. Cette structure génétique apparaît dès les fines échelles spatiales et suggère une dispersion par le vecteur limitée et un rôle important de la dérive génétique. De même, la présence d'individus avec des génotypes hybrides entre des échantillons distants géographiquement et l'assignation d'individus vers des échantillons éloignés de ceux dont ils sont issus géographiquement soulignent le rôle potentiel de l'homme dans la dispersion à longue distance du PWN. Cependant, un nombre d'échantillons plus important est nécessaire pour étudier préciser la dispersion longue distance du PWN.

Article n° 1

Mallez S, Castagnone C, Espada M, Vieira P, Eisenback JD, Mota M, Guillemaud T, Castagnone-Sereno P. (2013). First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. PLoS ONE 8(3): e59165. doi:10.1371/journal.pone.0059165.

First Insights into the Genetic Diversity of the Pinewood Nematode in Its Native Area Using New Polymorphic Microsatellite Loci

Sophie Mallez^{1,2,3,4*}, Chantal Castagnone^{2,3,4}, Margarida Espada⁵, Paulo Vieira⁵, Jonathan D. Eisenback⁶, Manuel Mota⁵, Thomas Guillemaud^{1,3,4}, Philippe Castagnone-Sereno^{2,3,4}

1 INRA, UMR 1355 Institut Sophia Agrobiotech, Equipe Biologie des Populations Introduites, Sophia Antipolis, France, **2** INRA, UMR 1355 Institut Sophia Agrobiotech, Equipe Interactions Plantes-Nématodes, Sophia Antipolis, France, **3** Université de Nice Sophia Antipolis, UMR Institut Sophia Agrobiotech, Sophia Antipolis, France, **4** CNRS, UMR 7254 Institut Sophia Agrobiotech, Sophia Antipolis, France, **5** NemaLab/ICAAM – Instituto de Ciências Agrárias e Ambientais Mediterrâneas, Universidade de Évora – Núcleo da Mitra, Évora, Portugal, **6** Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, Virginia, United States of America

Abstract

The pinewood nematode, *Bursaphelenchus xylophilus*, native to North America, is the causative agent of pine wilt disease and among the most important invasive forest pests in the East-Asian countries, such as Japan and China. Since 1999, it has been found in Europe in the Iberian Peninsula, where it also causes significant damage. In a previous study, 94 pairs of microsatellite primers have been identified *in silico* in the pinewood nematode genome. In the present study, specific PCR amplifications and polymorphism tests to validate these loci were performed and 17 microsatellite loci that were suitable for routine analysis of *B. xylophilus* genetic diversity were selected. The polymorphism of these markers was evaluated on nematodes from four field origins and one laboratory collection strain, all originate from the native area. The number of alleles and the expected heterozygosity varied between 2 and 11 and between 0.039 and 0.777, respectively. First insights into the population genetic structure of *B. xylophilus* were obtained using clustering and multivariate methods on the genotypes obtained from the field samples. The results showed that the pinewood nematode genetic diversity is spatially structured at the scale of the pine tree and probably at larger scales. The role of dispersal by the insect vector *versus* human activities in shaping this structure is discussed.

Citation: Mallez S, Castagnone C, Espada M, Vieira P, Eisenback JD, et al. (2013) First Insights into the Genetic Diversity of the Pinewood Nematode in Its Native Area Using New Polymorphic Microsatellite Loci. PLoS ONE 8(3): e59165. doi:10.1371/journal.pone.0059165

Editor: Adam Stow, Macquarie University, Australia

Received: December 12, 2012; **Accepted:** February 12, 2013; **Published:** March 15, 2013

Copyright: © 2013 Mallez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a grant from European Union (KBBE.2010.1.4-09 REPHRAME project). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sophie.mallez@sophia.inra.fr

Introduction

The pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Burher, 1934) Nickle, 1970 (Nematoda: Aphelenchoiidae) [1], native to North America and widely distributed in the USA and Canada [2], is an invasive pest of pine forests. At a local scale, it is usually transmitted by a cerambycid beetle, belonging to the genus *Monochamus* [3–6]. Probably due to international trade, including wood, wood products and shipping containers, the pinewood nematode reached other continents at the beginning of the twentieth century [7], where it successfully established and caused pine wilt disease. As a consequence it has become a serious threat to coniferous forests worldwide, especially pine forests [8]. The pine wilt disease is now considered the most serious disease of forest trees in East Asian countries, such as China, Taiwan, South Korea and Japan [9,10]. The presence of PWN was detected for the first time in Europe in 1999, in Peninsula of Setúbal in Portugal [11]. New outbreaks have been identified since 2008 in the center of Portugal and on Madeira Island in 2009 [12]; and more recently in Spain [13,14].

In light of the significant risks for European forests along with environmental, economic and social impacts [15], there is an

urgent need to develop effective pest management of PWN. In particular, it is critical to understand the invasion and colonization of this pest, including the risk of non-vectored spread of PWN to healthy forests. Several studies have already attempted to decipher the invasion routes of PWN including the detection of the source of invading populations [16–18] and the inference of the history of the outbreaks in Asia [19–22] and in Europe [23,24]. Although an American origin of Japanese populations is now widely accepted [25–27], many questions remain concerning the invasion routes of PWN. Various limitations in these studies can be invoked, including (i) a low number of available genetic markers that can be used on single PWN individual due to the small size of the nematode, (ii) the use of too few field samples from both the invaded and native areas, (iii) the use of culture collection samples instead of field samples and (iv) the lack of use of adequate statistical methods devoted to invasion route inference, as those presented by Estoup and Guillemaud [28].

Microsatellite markers are widely used in population genetics studies [29,30] and specifically in invasion route inference [28]. In a preliminary study, a PWN-specific microsatellite enriched genomic library was built and sequenced using high-throughput 454 GS-FLX Titanium pyrosequencing (Roche Diagnostics) [31].

In short, genomic DNA was obtained from a pool of thousands of PWN from a laboratory collection and enriched by hybridization in the following microsatellite motifs: [(AG)₁₀, (AC)₁₀, (AAC)₈, (AGG)₈, (ACG)₈, (AAG)₈, (ACAT)₆ and (ATCT)₆]. Pyrosequencing yielded 12,286 sequences. The QDD program [32] was used to select sequences containing microsatellites with desirable properties and to design PCR primers pairs. Ninety four primer pairs were designed on sequences longer than 80 bp containing perfect or imperfect microsatellites with at least five repeated motifs.

The objectives of the present work were (i) to test in the laboratory the 94 microsatellite markers developed *in silico* by Malausa et al. [31], (ii) to set up multiplexed PCR reactions of specific microsatellite markers for routine use in PWN and (iii) to use them to gain first insights into the PWN genetic diversity and structure in its native area, which constitutes a prerequisite for deciphering its worldwide invasion routes [28].

Materials and Methods

Biological material

No permission was required to collect samples of this species in the native area and we obtained an official agreement from the French authorities (#2012060-0004) to manipulate this quarantine organism in the Institute Sophia Agrobiotech. In this study, a total of 115 individuals grouped into field samples and collection strains from native (USA and Canada) and invasive (Japan, China and Portugal) areas, were used. The characteristics of the samples are listed in Table 1. The field samples came from the native area (USA) and were extracted from wood samples that were collected directly from field locations. Each field sample corresponds to a single tree and consisted of between 15 and 31 individuals of mixed life stages per tree. The trees from Nebraska were close to each other (less than 5 meters) and distant about 500 km from the Missouri trees, which were about 50 km from each other. Nematodes were extracted with a Baermann funnel [33]. The collection strains, both from native and invasive areas, came from cultures that have been reared in the INRA laboratory since 1986 for the oldest strain. This collection was derived from original isolates of about 500 nematodes and is maintained monoxenically on *Botrytis cinerea* (deBary) Whetzel at 15 °C. Individuals were stored in DESS [34] at 4 °C before DNA was extracted.

DNA extraction

Genomic DNA was extracted, as described hereafter, by thermal shock from single individuals [35]. Each individual was transferred to 18 µl of lysis buffer (Taq buffer with MgCl₂ 10X, Taq Core Kits10, MP Biomedicals; 60 mg.ml⁻¹ Proteinase K and sterile distilled H₂O) and was then put at -80 °C for 45 min, and immediately transferred to 60 °C for 60 min and finally to 95 °C for 15 min in a Biometra® T3-Thermoblock Thermocycler.

Microsatellite markers validation

To avoid the presence of null alleles that are common in microsatellite markers [36], we tested the PCR amplification of the 94 primer pairs designed by Malausa et al. [31] on 18 individuals from the collection strains (Table 1). This first step was carried out using the following procedure: PCR amplifications were performed in a final volume of 25 µl containing 2 µl of genomic DNA extracted as described above, 2.5 µl of Taq buffer with MgCl₂ (10X, Taq Core Kits 10, MP Biomedicals), 0.2 µl of Taq DNA Polymerase (5 U/µl, Taq Core Kits 10, MP Biomedicals), 1.2 µl of dNTPs (10 mM, Taq Core Kits 10, MP Biomedicals), 0.5 µl of each primer (10 µM, Eurogentec) and sterile distilled H₂O. The amplification reactions were performed in a T3-Thermoblock Thermocycler Biometra® and included a 10 min denaturation step at 95 °C, followed by 40 cycles of 30 sec at 95 °C, 30 sec at 55 °C and 1 min at 72 °C, followed by a final extension step at 72 °C for 10 min. The markers which gave positive PCR amplifications were then used in fluorescent PCR in order to analyze their polymorphism. This step was conducted on 100 individuals from the field samples and one single collection strain (US10) from the native area (Table 1). Two microsatellite markers from the literature, Bx07 and Bx08 [22], were also added at this step since they amplified well in our PCR conditions. PCR amplifications were performed in 10 µl containing 1X QIAGEN Multiplex Master Mix, 2 µM of each primer with forward primers labeled with a fluorescent dye (6-FAM, VIC, PET or NED) on the 5' end and 2 µl of genomic DNA extracted by thermal shock as explained above. The amplification reactions were performed in a Biometra® T3-Thermoblock Thermocycler and included a 15 min denaturation step at 95 °C, followed by 28 or 33 cycles (depending on the primer pairs, see Results) of 30 sec at 94 °C, 1.5 min at 55 °C, and 1 min at 72 °C, followed by a final extension step of 30 min at 60 °C. Genotype scoring was performed using an ABI 3700 sequencer (Applied Biosystems) with the 500 LIZ™ GeneScan™

Table 1. Characteristics of the samples of *Bursaphelenchus xylophilus* used in this study.

Type of samples	Code	No. individuals	Origin	Host tree
Field samples	MO1	31	USA - Missouri - Columbia	<i>Pinus sylvestris</i> L.
	MO2	23	USA - Missouri - Columbia	<i>P. sylvestris</i>
	NE1	16	USA - Nebraska - Davey	<i>P. sylvestris</i>
	NE2	15	USA - Nebraska - Davey	<i>P. sylvestris</i>
Collection strains	US10	15	USA - Minnesota	<i>Abies balsamea</i> (L.) Mill.
	US9	3	USA - Arizona - Tucson	<i>P. halepensis</i> Miller
	J10	3	Japan - Nishiaizu (Fukushima pref.)	<i>P. densiflora</i> Siebold & Zucc.
	Bx China	3	China	no information
	Bx Portugal	3	Portugal	<i>P. pinaster</i> Alton
	01.602.1	3	Intercepted on packaging wood from Canada	Packaging wood

doi:10.1371/journal.pone.0059165.t001

size standard (Applied Biosystems) and Genemarker™ version 1.75 software (SoftGenetics LLC).

Genetic diversity analyses

We computed observed (H_o) and expected (H_e) heterozygosities using Genetix 4.05 [37]. Deviations from Hardy-Weinberg equilibrium (HWE), for each locus and globally, were tested using Genepop version 4.1.3 [38]. Deviations from linkage equilibrium between loci were tested using the log likelihood ratio statistic in Genepop version 4.1.3 [38]. We took into account multiple testing (in the case of HWE tests) and non-independence between tests (in the case of linkage tests) by using the false discovery rate (FDR) correction [39] and the sequential Bonferroni adjustment [40], respectively. To quantify any inferred deviation from HWE, we calculated the Weir & Cockerham's estimate of F_{IS} [41] using Genepop [38]. Differences in mean allelic richness, computed using Fstat version 2.9.3.2 [42] and mean expected heterozygosity between field samples and collection strain (US10) were tested with the one-sided non-parametric test of Wilcoxon (with greater genetic diversity in the field samples), with the locus as the repeat unit, using R version 2.14.2 [43]. For the mean allelic richness analysis, the microsatellite marker PWN_35 was excluded because it presented missing data, which too greatly reduced the number of individuals taken into account to compute the allelic richness.

Genetic structure analyses

Nematodes from the field were sampled at the same time and reflect at least part of the genetic diversity existing in the field. This is not true for the collection strain, for which we do not have precise sampling information and for which many generations of genetic drift may have distorted the genotypic frequencies. Hence, only the 85 individuals from field origins were used for population genetics structure analyses. We first tested the hypothesis of genotypic frequencies homogeneity among samples using the exact test of Fisher [44] provided by Genepop version 4.1.3 [38]. Since multiple tests were performed, we adjusted the significance level using FDR correction [39]. We also performed an analysis of molecular variance (AMOVA), allowing measurement of the hierarchical distribution of the genetic variability, using Arlequin version 3.5.1.2 [45]. The different sources of variability tested were the following: within samples (i.e. trees), among samples within groups (i.e. states, Nebraska and Missouri) and among groups. The significance of the variance components associated with different levels of structure was tested by performing 20,000 permutations. We then studied the samples using the Bayesian assignment approach implemented in Structure version 2.3.4 [46]. This Bayesian method uses individual multilocus genotypes to infer clusters of individuals that minimize Hardy Weinberg and linkage disequilibria. An admixture model with correlated allele frequencies was used [47]. Ten independent runs for $K = 1$ to 12 were carried out each with a Markov Chain Monte Carlo (MCMC) of 150 000 iterations following a burn-in period of 50 000 iterations. Default values were maintained for all other parameters. The number of clusters was determined using the method of Evanno et al. [48]. Finally, we used a multivariate method, the Discriminant Analysis of Principal Components (DAPC) recently developed by Jombart et al. [49] because of its versatility. This method does not rely on any population genetics model and it is not constrained by any assumptions on HWE or linkage equilibrium. DAPC was performed using adegenet package [50] in R version 2.14.2 [43]. The number of clusters K varied from 1 to 12 and the number of inferred clusters was determined according to the Bayesian Information Criterion (BIC). The chosen number of clusters is

the minimum number of clusters after which the BIC increases or decreases by a negligible amount [50].

Results

Validation of microsatellite markers

Among the 94 microsatellite primer pairs designed by Malausa et al. [31], 25 gave positive PCR amplification at the predicted size for the 18 individuals (data not shown) and were consequently tested for their polymorphism. Fifteen out of these 25 markers and the two markers, Bx07 and Bx08, of Zhou et al. [22] could be unambiguously scored and were polymorphic (Tables 2 and 3). These 17 markers could be amplified in three PCR multiplex reactions: MA28 and MB28 multiplex panels, with 28 cycles in PCR amplification and MC33 multiplex panel with 33 cycles in PCR amplification (Table 2). The 15 markers developed here have been deposited in EMBL-Bank and accession numbers are shown in Table 2.

Genetic diversity

The total number of alleles per locus over all samples varied from 2 to 11, with a mean of 4.7. In Missouri samples, more than four microsatellite markers displayed five alleles or more. The expected heterozygosity per locus over all samples ranged from 0.142 to 0.825 (Table 3). Deviations from HWE associated with heterozygote deficiency were detected in 5 loci (PWN_51, PWN_60, PWN_62, PWN_79 and Bx07; Table 3). Significant linkage disequilibrium was detected between markers PWN_51 and Bx07 after sequential Bonferroni adjustment. In addition all alleles of marker PWN_51 have a length that is exactly 63 bp shorter than those of Bx_07.

The mean number of alleles per sample ranged from 1.1 to 3.5. The mean allelic richness and the mean expected heterozygosity were between 1.49 and 3.13 and between 0.144 and 0.385, respectively (Table 3). The mean allelic richness was generally larger in field samples than in the collection strain, with a reduction of 22% to 52% in the collection strain depending on the field samples. The observed differences were significant for MO1 (Wilcoxon's test, $p = 0.022$) and MO2 ($p = 0.020$) and non-significant for NE1 and NE2 (Wilcoxon's tests, $p > 0.1$). Expected heterozygosities were also lower in the collection strain with a reduction of 31% to 63% compared to the field samples. Only two significant (Wilcoxon's test, MO1, $p = 0.018$ and MO2, $p = 0.017$) larger mean expected heterozygosities were found in field samples compared to collection strain. Over all loci, HWE was rejected in 3 samples: MO1, MO2 and NE2 (Fisher's exact tests, $p < 10^{-3}$).

Assessing the genetic structure

Nematodes from the four field origins were significantly differentiated either within or between states (Fisher's exact tests, $p < 10^{-3}$). Moreover, the analysis of molecular variance revealed that the majority of the genetic variance was explained by the variation between individuals within trees (75.27%) and that the proportion of variance was much more important between states (16.64%) than between trees within states (8.09%).

Results of the clustering method using Structure and the multivariate method using DAPC are visualized and summarized in Figure 1. Both methods suggest the existence of three clusters. These three clusters were supported by a mean Structure co-ancestry coefficient larger than 93%. In DAPC, all individuals were assigned to the three clusters so that no 'ghost' population was inferred. The three clusters inferred by the two different methods were very similar, with one identical cluster and with only four individuals (NE1-4; NE2-6; NE2-8 and NE2-9) assigned

Table 2. Characteristics of 15 microsatellite markers developed for *Bursaphelenchus xylophilus*.

Locus	Primer sequence (5' - 3')	Motif repeat	Fluorescent label/ Multiplex panel	Allele size range (bp)	Accession Number
PWN_3	F : GAAATCTGGGAGCAAAACA R : ACCGCACTCGCACTTAGATT	(CT) ₈	6FAM/MC33	215–227	HF563643
PWN_6	F : GGAATTAGGCGTCCACAAGA R : TGCTGTATAACATTGCTCTTCG	(AG) ₇	6FAM/MC33	126–131	HF563644
PWN_26	F : GAAAACTTAGGCTGGGGGA R : TAGTGACGACTCATCCGCTG	(TG) ₅	PET/MA28	157–160	HF563645
PWN_30	F : ACCTAGCGTCGAAACCTT R : ATAGCAGCAGGTCAAATCCG	(TG) ₅	VIC/MB28	207–209	HF563646
PWN_34	F : CCATTGCCCAAAGGATTA R : ACGTAGCATTGGAGTGACC	(CT) ₇	PET/MC33	82–95	HF563647
PWN_35	F : ACCGCCTGGTAACCGAGT R : TTGGACACTGCGAGTAAGGA	(GA) ₆	PET/MB28	185–193	HF563648
PWN_49	F : CTGGGAGTTCTTTTGCTCG R : GCAACAATCGTTAGTGGA	(AAC) ₅	PET/MA28	174–177	HF563649
PWN_51	F : GGAAGAGACTTGACCCGAAA R : GGAAAAGAGTCTCACGTCAA	(AG) ₇	6FAM/MC33	84–96	HF563650
PWN_54	F : ACCTTCACACTGTAGCCGC R : CCGGTCATCATAATCTCTGATCT	(AG) ₇	PET/MA28	113–119	HF563651
PWN_56	F : TCTTCACATTAATCTTGCTCC R : AACGATTAGGAACGAGTGG	(CA) ₈	PET/MC33	185–195	HF563652
PWN_60	F : GCGGAAACGGATAAAGGAAT R : TTCTTCCCAACCTTCTCC	(CA) ₉	VIC/MB28	129–147	HF563653
PWN_62	F : GAGCTATAGCCCTGCCTTT R : AGCCTTGCGAAGAAACAAA	(CT) ₆	6FAM/MA28	112–124	HF563654
PWN_79	F : TGGATACAAACGGTTGAGGA R : AACCTCATCTGTCCGTGGAT	G(GA) ₂ G(GA) ₈ A(T) ₅	NED/MB28	107–114	HF563655
PWN_80	F : AATTGGTGCTCTGTATGGC R : CGGCTTACTCTTTGTCCCA	TG(TGT) ₅ TG	VIC/MB28	78–88	HF563656
PWN_84	F : CCGTGTTTCAACTCATTC R : TTTGATCCGATTACCTTCGG	(CT) ₂ T(CT) ₅ C	PET/MC33	129–137	HF563657
Bx07	F : AACGAAAAGAGTCTCACG R : TAGGCCCTCCTTGACAAAAGC	(TC) ₁₀	6FAM/MB28	146–157	[22]
Bx08	F : CTGCCTATTTTCGACTCTC R : CAAGGATCGTGTCCTCTTTTG	(CT) ₁₀	NED/MA28	105–113	[22]

Characteristics of the microsatellite markers from Zhou et al. [22], Bx07 and Bx08, are also given.
doi:10.1371/journal.pone.0059165.t002

differently to the two remaining clusters. One of the clusters consisted of individuals from Nebraska and the two others were shared between the two samples from Missouri, mixing individuals from different trees. The four individuals from Nebraska mentioned above were assigned either to the Nebraska's cluster or to one of the two clusters found in Missouri, depending on the method used (see Figure 1).

Discussion

We developed 15 new microsatellite markers with the following properties: (i) they are easily usable in routine conditions, (ii) they can be used together on single individuals, (iii) they are polymorphic at the individual level and (iv) only three multiplex PCR reactions are necessary to genotype each PWN individual.

Two markers, Bx07 and Bx08, from Zhou et al. [22], were also added in two multiplex PCR reactions, because they are polymorphic and they amplified well in our PCR conditions. Nonetheless, it is important to note that we observed significant linkage disequilibrium between Bx07 and PWN_51 and all alleles of Bx07 are exactly 63 bp larger than the corresponding alleles of PWN_51. This leads us to suppose that these two markers correspond to the same locus. However, we could not verify this hypothesis because the sequence of Bx07 is not yet publicly available. Several markers (PWN_51, PWN_60, PWN_62, PWN_79 and Bx07) exhibited also a significant heterozygote deficiency that may be due to two reasons: (i) the presence of null alleles, i.e., alleles that are not amplified by the multiplex PCR, and (ii) a Wahlund effect, i.e., a direct consequence of the existence of subdivisions in the studied populations. As no individuals with

Table 3. Summary of standard population genetics analyses for each sample.

Locus	Na (total)	He (total)	MO1 (n = 31)				MO2 (n = 23)				NE1 (n = 16)				NE2 (n = 15)				US10 (n = 13)			
			Na	Ho	He	Fis	Na	Ho	He	Fis	Na	Ho	He	Fis	Na	Ho	He	Fis	Na	Ho	He	Fis
PWN_3	3	0.309	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—	2	0.692	0.471	—0.500
PWN_6	3	0.291	2	0.194	0.178	−0.091	2	0.044	0.044	0	1	—	—	—	1	—	—	—	1	—	—	—
PWN_26	3	0.306	2	0.039	0.039	0	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—
PWN_30	2	0.283	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—
PWN_34	5	0.305	2	0.065	0.064	−0.017	4	0.174	0.205	0.154	2	0.063	0.063	0	1	—	—	—	2	0.077	0.323	0.769
PWN_35	5	0.735	3	0.500	0.668	0.256	3	0.462	0.655	0.304	1	—	—	—	1	—	—	—	2	0.154	0.148	−0.044
PWN_49	2	0.253	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—	1	—	—	—
PWN_51 α	6	0.730	5	0.484	0.649	0.258*	6	0.636	0.771	0.178	2	0.563	0.417	−0.364	3	0.400	0.690	0.429*	2	0.154	0.148	−0.044
PWN_54	7	0.639	4	0.552	0.571	0.035	6	0.476	0.617	0.232	4	0.286	0.325	0.126	2	0.083	0.083	0	2	0.692	0.508	−0.385
PWN_56	4	0.659	4	0.400	0.432	0.076	4	0.364	0.648	0.445	3	0.313	0.280	−0.119	2	0.308	0.443	0.314	2	0.231	0.212	−0.091
PWN_60	11	0.825	8	0.467	0.746	0.378*	6	0.381	0.612	0.384*	2	0.133	0.129	−0.037	3	0.643	0.680	0.057	1	—	—	—
PWN_62	9	0.644	7	0.484	0.672	0.283	7	0.381	0.702	0.463*	4	0.500	0.730	0.322	3	0.214	0.519	0.596*	2	0.077	0.077	0.000
PWN_79	6	0.822	5	0.733	0.777	0.057*	5	0.619	0.741	0.168	3	0.467	0.522	0.109	4	0.571	0.632	0.010	1	—	—	—
PWN_80	2	0.142	2	0.065	0.064	−0.017	2	0.174	0.162	−0.073	2	0.125	0.121	−0.035	1	—	—	—	2	0.539	0.409	−0.333
PWN_84	4	0.266	2	0.032	0.094	0.659	1	—	—	—	2	0.063	0.063	0	1	—	—	—	1	—	—	—
Bx07 α	6	0.729	5	0.484	0.649	0.258	6	0.600	0.774	0.230	2	0.500	0.387	−0.304	3	0.250	0.692	0.649*	2	0.154	0.148	−0.044
Bx08	5	0.593	3	0.323	0.349	0.076	4	0.524	0.614	0.151	2	0.385	0.508	0.250	3	0.333	0.297	−0.129	1	—	—	—
All loci	4.9	0.482	3.4	0.284	0.350	0.192*	3.5	0.284	0.385	0.264*	2.0	0.199	0.209	0.036	1.9	0.165	0.237	0.306*	1.1	0.163	0.144	−0.140

Note: Na (total), Na, Ho and He refer to as the total number of alleles per locus over all samples, the number of alleles per locus in each sample, the observed heterozygosity and the expected heterozygosity, respectively. Fis was calculated after Weir & Cockerham [41]. The last row gives mean numbers of alleles, mean heterozygosities and Fis calculated over all loci. '*' indicates that the HWE test is significant after FDR correction [39] (except for the last row). ' α ' indicates the microsatellite markers involved in significant linkage disequilibria after sequential Bonferroni adjustment [40]. '—' means that for monomorphic markers, Ho, He and Fis were not computed.

doi:10.1371/journal.pone.0059165.t003

missing genotypes were observed at these loci, the first hypothesis is not likely. Moreover, clustering analyses corroborated the second hypothesis. Using these markers, the level of polymorphism detected in our samples was low to moderate but sufficient for population genetics analyses. Moreover, it lies in the range

typically found in the literature for other phytoparasitic nematode species [51–56]. This new set of markers provides a useful tool, appropriate to implement analytical methods in population genetics due to its resolution at the individual level. Specifically, these markers will be very useful in identifying the source of

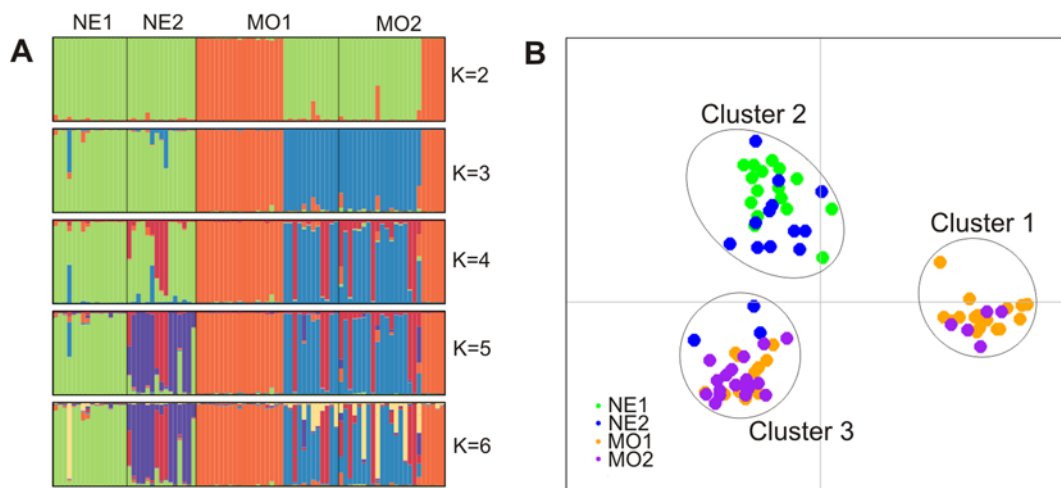


Figure 1. Genetic structure of the PWN field samples from the USA. A, Barplots of Structure of the coefficient of co-ancestry for K=2, 3, 4, 5 and 6 clusters. Each bar corresponds to one individual nematode and each cluster is represented by a color. The number of clusters inferred was K=3, based on the ΔK of Evanno et al. [48]. B, DAPC scatterplot showing the first two principal components of the DAPC for K=3, the number of cluster being inferred from the Bayesian Information Criterion (BIC).

doi:10.1371/journal.pone.0059165.g001

invasive outbreaks and in deciphering the invasion routes of the PWN. Practically, they will be used to obtain multilocus genotypes (MLGs) of numerous samples from the native and invaded areas. These MLGs will be analyzed using recent and appropriate methods devoted to this question, like approximate Bayesian computation [28,57] which has already been successfully used in other invading species [58–60]. Until now, most markers developed for the PWN, such as AFLP [19,21], RAPD [16,24], ISSR [16,18], IGS [18], homologous DNA probes [17] or cytochrome b and cellulase gene sequences [23] suffer from technical limitations. They either displayed a low level of polymorphism [23], were not codominant [16,18,19,21,24] and/or were used on pooled collections of individuals [16–19,21,23,24] with results that are difficult to interpret. A few microsatellite markers have already been developed for the PWN [20,22] but they either required the pre-amplification of genomic DNA [20], which can cause an artifactual polymorphism, or were too few to obtain a clear image of the genetic diversity of the samples analyzed [22]. The reason for these problems is that, in general, the minute quantity of DNA from each individual is an obstacle to obtain individual multilocus genotypes at a large number of markers. Multiplex PCR reactions, in addition to reducing costs, allow this problem to be overcome by amplifying several microsatellite markers in a single PCR reaction ([61] for a review, [62]). The main advantage of the tools developed here is thus the possibility to genotype individuals using three multiplex PCR reactions. Only three PCR reactions per individual are needed to obtain the diploid multilocus genotypes of 17 microsatellite markers without any DNA pre-amplification step.

We further used the 17 microsatellite markers on 100 individuals PWN collected from four field locations and one laboratory collection strain to obtain first insights into the genetic variability of *B. xylophilus* at the individual level in its native area.

First, the number of alleles and the expected heterozygosity were lower in the collection strain than in field samples with large reductions of 22–52% and of 31–63%, respectively. Although the statistical significance of this result was not clear, it seems that the collection strain has likely suffered a loss of genetic diversity compared to field samples, particularly the Missouri samples. This difference between laboratory strains and wild populations/field samples has already been explored and demonstrated in other species. For instance, Kim et al. [63] observed a loss of 15–39% genetic diversity in the non-diapause colony of the western corn rootworm (*Diabrotica virgifera virgifera* Le Conte, 1868) compared with contemporary wild populations, depending on the parameter measured. Similar results were obtained by Coe et al. [64] with the zebrafish (*Danio rerio* Hamilton, 1863) with the allelic richness for all four strains less than 20% of that found in the wild fish. The maintenance of strains in laboratory collections and, in our specific case, the very short generation time, i.e. ± 5 –7 days at 20 °C [65], and the need to transfer nematodes from one Petri dish to another to ensure the viability of the strain, probably create recurrent bottlenecks, which in turn decrease the genetic diversity. Further investigations are needed to clarify this observation. However, in any case, this potential loss of genetic diversity is important to take into account when performing population genetic diversity analyses using medium and long-term culture collection samples.

Second, intra-sample variability was near HWE and linkage equilibrium. This result confirms that the PWN reproduces sexually *in natura* with no evidence of deviation from random mating between individuals. Moreover, an important part of the genetic variance detected here corresponded to inter-individual variation within a tree. This suggests that several nematodes enter

into the tree and more specifically that several nematodes effectively reproduce and contribute to the growth of the population inside the tree. This observation is supported by the presence of more than 4 alleles per locus for some microsatellite markers at the tree level. Different results were observed by Zhou et al. [22] in Japan : (i) a very low genetic diversity was detected at the tree level, with 418 individuals (out of 420) presenting the same individual genotype on 14 trees sampled ; (ii) the genetic variability was more important between than within trees. This difference can be explained by (i) a technical limitation, owing to the small number of markers (only four) used in this study leading to a large variance of the statistics summarizing genetic variation; and by (ii) the loss of genetic diversity that generally occurs during invasion, resulting in lower genetic diversity in invasive populations than in native ones [66,67].

Finally, the various samples displayed significant genetic differences, highlighting the existence of a spatial genetic structure. Spatial differentiation exists at very short scale, with neighboring trees of Nebraska significantly differentiated. This suggests that the PWN dispersal, whether active or passive, can be spatially limited even at a short scale and that genetic drift may play an important role. Furthermore, both methods used in this study (Bayesian assignment and multivariate methods) inferred three clusters among the field samples analyzed. Each cluster consisted of individuals from different trees, reinforcing the existence of a genetic structure within and between trees. Different clusters were identified within trees scale (Missouri trees) suggesting that different beetles carrying genetically differentiated nematode populations infected a single tree. The individuals from the Nebraska trees, close to each other, were grouped in a single cluster. In addition, both Missouri trees exhibited the same two genetic clusters. These local genetic similarities probably result from efficient short distance dispersal mediated by the insect vector [68]. Some nematodes from Nebraska were also assigned to a cluster mainly formed by Missouri individuals (results of DAPC method) or presented hybrid genotypes between Nebraska and Missouri clusters (results of Bayesian method) despite the large geographical distance between them (more than 500 km). This result is an agreement with the potentially important role of the human-induced dispersal, already proposed in others studies on the PWN [69,70] and in other nematode species [71]. However, too few samples were used in this study to provide clear evidence of long-distance dispersal. With that respect, a hierarchical sampling scheme with nematodes sampled from various trees located in different groups of trees situated in different forests should be implemented. This would allow a precise assessment of the population genetic structure of the PWN to better determine the spatial range of nematode dispersal and the scale at which populations function.

Acknowledgments

We would like to thank Aurélie Blin and Pedro Barbosa for helpful technical support, Douglas LeDoux and Julia Thompson (Missouri Department of Agriculture, USA), and Mark Harrell (Nebraska Forest Service, USA) for supplying the field isolates of *B. xylophilus* from Missouri and Nebraska, respectively.

Author Contributions

Performed the sampling: ME PV JE MM. Performed the experiments: SM CC ME PV JE. Analyzed the data: SM TG PCS. Contributed reagents/materials/analysis tools: CC JE MM. Wrote the paper: SM TG PCS.

References

- Hunt DJ (2008) A checklist of the Aphelenchoidea (Nematoda: Tylenchina). *Journal of Nematode Morphology and Systematics* 10: 99–135.
- Dropkin VH, Foudin A, Kondo E, Linit MJ, Smith M, et al. (1981) Pinewood nematode: a threat to US forests? *Plant Disease* 65: 1022–1027.
- Linit MJ (1988) Nematode - vector relationships in the pine wilt disease system. *Journal of Nematology* 20: 227–235.
- Mamiya Y (1972) Pine wood nematode *Bursaphelenchus lignicolus* Mamiya and Kiyohara, as a causal agent of pine wilting disease. *Review of Plant Protection Research* 5: 46–60.
- Naves PM, Camacho S, de Sousa EM, Quartau JA (2007) Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through feeding activity of *Monochamus galloprovincialis* (Col., Cerambycidae). *Journal of Applied Entomology* 131: 21–25.
- Sousa E, Bravo MA, Pires J, Naves P, Penas AC, et al. (2001) *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal. *Nematology* 3: 89–91.
- Mamiya Y (1988) History of pine wilt disease in Japan. *Journal of Nematology* 20: 219–226.
- Suzuki K (2002) Pine wilt disease - a threat to pine forest in Europe. *Dendrobiology* 48: 71–74.
- Mamiya Y (1976) Pine wilt disease caused by the pine wood nematode, *Bursaphelenchus lignicolus*, in Japan. *Japan Agricultural Research Quarterly* 10: 206–211.
- Mamiya Y (1983) Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 21: 201–220.
- Mota M, Braasch H, Bravo MA, Penas AC, Burgermeister W, et al. (1999) First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1: 727–734.
- Fonseca L, Cardoso JMS, Lopes A, Pestana M, Abreu F, et al. (2012) The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia* 49: 96–103.
- Abelleira A, Picoaga A, Mansilla JP, Aguín O (2011) Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant Disease* 95: 776.
- Robertson L, Cobacho Arcos S, Escuer M, Santiago Merino R, Esparrago G, et al. (2011) Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhrer, 1934 (Nickle, 1970) in Spain. *Nematology* 13: 755–757.
- Soliman T, Mourits MCM, van der Werf W, Hengeveld GM, Robinet C, et al. (2012) Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *PLoS ONE* 7(9): e45505. doi:10.1371/journal.pone.0045505.
- Metge K, Burgermeister W (2008) Analysis of *Bursaphelenchus xylophilus* (Nematoda: Parasitaphelenchidae) provenances using ISSR and RAPD fingerprints. In: Mota M, Vieira P, editors. *Pine wilt disease: a worldwide threat to forest ecosystems*. pp. 175–186.
- Tares S, Abad P, Bruguié N, de Guiran G (1992) Identification and evidence for relationships among geographical isolates of *Bursaphelenchus* spp. (pinewood nematode) using homologous DNA probes. *Heredity* 68: 157–164.
- Valadas V, Barbosa P, Espada M, Oliveira S, Mota M (2012) The pinewood nematode, *Bursaphelenchus xylophilus*, in Portugal: possible introductions and spread routes of a serious biological invasion revealed by molecular methods. *Nematology* 00: 1–13.
- Cheng XY, Cheng FX, Xu RM, Xie BY (2008) Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: Aphelenchoididae) and its possible spread routes in China. *Heredity* 100: 356–365.
- Jung J, Han H, Ryu SH, Kim W (2010) Microsatellite variation in the pinewood nematode, *Bursaphelenchus xylophilus* (Steiner and Buhrer) Nickle in South Korea. *Genes & Genomics* 32: 151–158.
- Jung J, Han H, Ryu SH, Kim W (2010) Amplified fragment length polymorphism analysis and genetic variation of the pinewood nematode *Bursaphelenchus xylophilus* in South Korea. *Animal Cells and Systems* 14: 31–36.
- Zhou ZH, Sakaue D, Wu BY, Hogetsu T (2007) Genetic structure of populations of the pinewood nematode *Bursaphelenchus xylophilus*, the pathogen of pine wilt disease, between and within pine forests. *Phytopathology* 97: 304–310.
- Valadas V, Laranjo M, Mota M, Oliveira S (2012) Molecular characterization of Portuguese populations of the pinewood nematode *Bursaphelenchus xylophilus* using cytochrome b and cellulase genes. *Journal of Helminthology*: 1–10. doi: 10.1017/S0022149X12000673
- Vieira P, Burgermeister W, Mota M, Metge K, Silva G (2007) Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses. *Journal of Nematology* 39: 118–126.
- Jones JT, Moens M, Mota M, Li HM, Kikuchi T (2008) *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host-parasite interactions. *Molecular Plant Pathology* 9: 357–368.
- Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, et al. (2011) Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog* 7(9): e1002219. doi:10.1371/journal.ppat.1002219.
- Wingfield MJ, Blanchette RA, Nicholls TH, Robbins K (1982) The pine wood nematode: a comparison of the situation in the United States and Japan. *Canadian Journal of Forest Research* 12: 71–75.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* 19: 4113–4130.
- Luikart G, England PR (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology & Evolution* 14: 253–256.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629.
- Malasa T, Gilles A, Meglecz E, Blanquart H, Duthoy S, et al. (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.
- Meglecz E, Costedoat C, Dubut V, Gilles A, Malasa T, et al. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26: 403–404.
- Viglierchio DR, Schmitt RV (1983) On the methodology of nematode extraction from field samples: Baermann funnel modifications. *Journal of Nematology* 15: 438–444.
- Yoder M, Tandingan De Ley I, King IW, Mundo-Ocampo M, Mann J, et al. (2006) DESS: a versatile solution for preserving morphology and extractable DNA of nematodes. *Nematology* 8: 367–376.
- Castagnone C, Abad P, Castagnone-Sereno P (2005) Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae). *European Journal of Plant Pathology* 112: 191–193.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93: 504–509.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier Montpellier (France).
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57: 289–300.
- Sokal RR, Rohlf FJ (1995) *Biometry. The Principles and Practice of Statistics in Biological Research*. Third Edition edn. W.H. Freeman and Company, New York. pp 236–242.
- Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
- Goudet J (2002) Fstat, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). : Available online <http://www2.unil.ch/popgen/softwares/fstat.htm>
- R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* 49: 1280–1283.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- He Y, Li HM, Brown DJF, Lamberti F, Moens M (2003) Isolation and characterisation of microsatellites for *Xiphinema* index using degenerate oligonucleotide primed PCR. *Nematology* 5: 809–819.
- Leach M, Agudelo P, Lawton-Rauh A (2012) Genetic variability of *Rotylenchulus reniformis*. *Plant Disease* 96: 30–36.
- Mulet K, Fargette M, Richaud M, Genson G, Castagnone-Sereno P (2011) Isolation of microsatellites from an enriched genomic library of the plant-parasitic nematode *Meloidogyne incognita* and their detection in other root-knot nematode species. *European Journal of Plant Pathology* 129: 501–505.
- Picard D, Plantard O, Scurrah M, Mugniéry D (2004) Inbreeding and population structure of the potato cyst nematode (*Globodera pallida*) in its native area (Peru). *Molecular Ecology* 13: 2899–2908.
- Plantard O, Porte C (2003) Isolation and characterization of microsatellite loci in the sugar beet cyst nematode *Heterodera schachtii*. *Molecular Ecology Notes* 3: 139–141.
- Thiéry M, Mugniéry D (2000) Microsatellite loci in the phytoparasitic nematode *Globodera*. *Genome* 43: 160–165.

57. Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162: 2025–2035.
58. Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, et al. (2010) Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *PLoS ONE* 5(3): e9743. doi:10.1371/journal.pone.0009743
59. Lombaert E, Guillemaud T, Thomas CE, Handley IJL, Li J, et al. (2011) Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird *Harmonia axyridis*. *Molecular Ecology* 20: 4654–4670.
60. Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, et al. (2005) Multiple transatlantic introductions of the western corn rootworm. *Science* 310: 992–992.
61. Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉGER P, et al. (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11: 591–611.
62. Karaiskou N, Primmer C (2008) PCR multiplexing for maximising genetic analyses with limited DNA samples: an example in the collared flycatcher, *Ficedula albicollis*. *Annales Zoologici Fennici* 45: 478–482.
63. Kim KS, French BW, Sumerford DV, Sappington TW (2007) Genetic diversity in laboratory colonies of western corn rootworm (Coleoptera : Chrysomelidae), including a nondiapause colony. *Environmental Entomology* 36: 637–645.
64. Coe TS, Hamilton PB, Griffiths AM, Hodgson DJ, Wahab MA, et al. (2009) Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies. *Ecotoxicology* 18: 144–150.
65. Mamiya Y (1975) The life history of the pine wood nematode, *Bursaphelenchus lignicolus*. *Japanese Journal of Nematology* 5: 16–25.
66. Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, et al. (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics* 32: 305–332.
67. Wares JP, Hughes AR, Grosberg RK (2005) Mechanisms that drive evolutionary change. In: Sax DF, Stachowicz JJ, Gaines SD, editors. *Species Invasions: Insights into Ecology, Evolution and Biogeography*. Sunderland, MA, USA: Sinauer Associates Inc. pp. 229–257.
68. Shibata E (1986) Dispersal movement of the adult Japanese pine sawyer, *Monochamus alternatus* Hope (Coleoptera, Cerambycidae) in a young pine forest. *Applied Entomology and Zoology* 21: 184–186.
69. Evans HF, McNamara DG, Braasch H, Chadœuf J, Magnusson C (1996) Pest Risk Analysis (PRA) for the territories of European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *OEPP/EPPO Bulletin* 26: 199–249.
70. Robinet C, Roques A, Pan HY, Fang GF, Ye JR, et al. (2009) Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS ONE* 4(2): e4646. doi:10.1371/journal.pone.0004646.
71. Plantard O, Porte C (2004) Population genetic structure of the sugar beet cyst nematode *Heterodera schachtii*: a gonochoristic and amphimictic species with highly inbred but weakly differentiated populations. *Molecular Ecology* 13: 33–41.

b. Diversité et structure génétiques en lien avec l'invasion

❖ Contexte et objectifs

Étudier les premières étapes des invasions biologiques est crucial pour améliorer les moyens de lutte. Il est en effet plus efficace d'agir dès les premières étapes des invasions (en termes de coûts et de temps notamment) que de tenter d'éradiquer l'envahisseur une fois que celui-ci est installé (Allendorf & Lundquist 2003; Simberloff et al. 2013). Identifier les routes d'invasion constitue une étape clé de ces premières étapes, comme explicité dans la partie I.3.a. Dans le cas précis du PWN, plusieurs études ont déjà essayé de déterminer la source des différents foyers envahissants (Cheng et al. 2008; Figueiredo et al. 2013; Fonseca et al. 2012; Metge & Burgermeister 2008; Pereira et al. 2013; Tares et al. 1992; Valadas et al. 2012; Valadas et al. 2012; Vieira et al. 2007; Zhang et al. 2008; Zhou et al. 2007). Ainsi, certaines études suggèrent une origine asiatique pour les populations portugaises (Figueiredo et al. 2013; Fonseca et al. 2012; Metge & Burgermeister 2008; Valadas et al. 2012; Vieira et al. 2007). Néanmoins comme préalablement abordé (voir partie III.1.a), ces études souffrent de limites et donc aucune étude à l'heure actuelle n'a encore clairement étudié les routes d'invasion du PWN avec des marqueurs moléculaires adaptés, des échantillons appropriés et des approches de génétique des populations adéquates (voir Estoup & Guillemaud 2010).

Ainsi, les objectifs de cette étude étaient de caractériser génétiquement les zones envahies pour déterminer le nombre d'événements d'introduction le plus probable et d'étudier les relations entre la zone native et les zones envahies, en tentant notamment de déterminer la (les) source(s) des populations européennes. De plus, au-delà des routes d'invasion, nous avons cherché à confirmer l'existence d'une forte structure génétique dans l'aire native du PWN (déjà suggérée lors de la mise au point des marqueurs microsatellites), profitant d'un nombre d'échantillons plus important.

❖ Matériel et méthodes

Pour cette étude, nous disposons d'échantillons de populations naturelles de trois zones géographiques : une partie de l'aire native, les USA avec 18 échantillons comptabilisant 391 individus et deux zones envahies, le Japon avec 7 échantillons et 210 individus et le Portugal/Madère avec 9 échantillons et 169 individus. Chaque individu a été génotypé à l'aide de marqueurs microsatellites développés précédemment et des analyses intra- et inter- zones ont été réalisées.

Nous avons ainsi évalué la diversité (nombre d'allèles et hétérozygotie attendue) et la structure (par analyse Bayésienne de groupement des individus, Pritchard et al. 2000) génétiques au sein de chaque zone. Pour l'aire native, l'échantillonnage effectué nous a également permis de tester l'existence d'un isolement par la distance (test de Mantel entre distances génétique et géographique). Nous avons ensuite tenté de déterminer la source la plus probable, dans l'aire native ou en Asie, des populations européennes par différentes méthodes : (i) mesures de F_{st} entre paires de populations (Weir & Cockerham 1984), (ii) vraisemblances d'assignation individuelle (Paetkau et al. 2004; Rannala & Mountain 1997), (iii) arbre de distances génétiques (Cavalli-Sforza & Edwards 1967; Saitou & Nei 1987) et (iv) analyse Bayésienne de groupement des individus (Pritchard et al. 2000).

❖ Résultats

Nous avons observé une diversité génétique faible à modérée aux USA et extrêmement faible dans les 2 zones envahies étudiées ici due à la présence de nombreux marqueurs monomorphes (10 pour le Japon et 15 pour le Portugal, sur 16). Nous avons ensuite mis en évidence une structure génétique spatiale importante aux USA avec présence d'isolement par la distance. En effet, tous les échantillons sont apparus significativement différenciés entre eux avec des valeurs de F_{st} pouvant être élevées (voir Tableau S1 dans

l'article). L'analyse de groupement des individus a mis en évidence que la plupart des échantillons s'individualisaient (voir Figure 2A dans l'article). De manière similaire, le Japon a présenté une structure génétique spatiale avec des échantillons significativement différenciés entre eux et des valeurs de F_{st} très élevées (à l'exception des 3 échantillons complètement identiques et présentant le même génotype multi-locus, voir Tableau S1 dans l'article), ainsi qu'un groupement des individus à l'échelle de l'échantillon (voir Figure 2B). Finalement, en Europe, 8 échantillons sur les 9 analysés sont apparus complètement identiques avec 162 individus (sur les 169) présentant le même génotype multi-locus. L'échantillon restant est apparu significativement différencié des autres (valeurs de F_{st} présentées dans le Tableau S1 dans l'article). L'analyse de groupement des individus a suggéré l'existence d'un seul groupe génétique à l'échelle du Portugal et de l'île de Madère (voir Figure 2C dans l'article). Les différentes méthodes utilisées pour tenter de déterminer l'origine des populations européennes ont donné des résultats contradictoires. Les mesures de F_{st} et les vraisemblances d'assignation individuelle ont suggéré une origine américaine pour les échantillons du Portugal et de l'île de Madère. À l'inverse, l'arbre de distances génétiques et l'analyse Bayésienne de groupement des individus ont suggéré une origine asiatique.

❖ Discussion

Nous avons obtenu trois résultats majeurs dans cette étude : la structuration importante dans l'aire native, le faible niveau de diversité génétique dans les zones envahies et des méthodes classiques et Bayésienne non concluantes pour déterminer l'origine des populations européennes.

Tout d'abord, nous avons confirmé la forte structuration génétique spatiale existante dans l'aire native et déjà suggérée dans l'étude précédente. Cela appuie d'autant plus l'idée que la dérive génétique est importante et qu'elle n'est pas compensée par une dispersion

efficace. La dispersion limitée du PWN peut refléter la complexité du processus de dispersion naturelle, qui est principalement dépendante de l'insecte vecteur (Akbulut & Stamps 2012). De plus, le PWN doit surmonter plusieurs obstacles pour réussir à se reproduire dans l'arbre (Futai 2013). Il doit notamment réussir à entrer dans l'insecte avant son émergence et à entrer dans l'arbre pendant que l'insecte se nourrit. Ces obstacles peuvent conduire à diminuer l'efficacité de la transmission du PWN (Kobayashi et al. 1984; Togashi 1985) et contribuer à la forte dérive génétique observée ici.

Ensuite, nous avons mis en évidence une diversité génétique très faible voire complètement nulle dans les zones envahies étudiées ici. Ce résultat peut être expliqué par la faible diversité intra-échantillon observée dans l'aire native et par des effets fondateurs/goulots d'étranglement pouvant avoir lieu lors de l'introduction d'organismes dans une nouvelle aire géographique (Allendorf & Lundquist 2003; Sakai et al. 2001). Il suggère fortement qu'un seul évènement d'introduction a eu lieu dans chacune des zones envahies et que chaque population introduite avait une petite taille efficace. La quasi-absence de diversité en Europe suggère fortement aussi que le deuxième foyer envahissant portugais correspond à l'expansion du premier et que le Portugal continental est à l'origine des populations retrouvées sur l'île de Madère.

Enfin, nous n'avons pas pu identifier l'origine des populations européennes avec les méthodes d'analyses classiques et Bayésienne utilisées dans cette étude puisqu'elles ont donné des résultats contradictoires. Plusieurs hypothèses ont été avancées pour expliquer cela : (i) le manque de diversité génétique, ayant pu conduire à un manque de puissance discriminatrice, (ii) la forte structuration génétique de l'aire native, nécessitant un échantillonnage quasi exhaustif – que nous n'avons pas – pour saisir la variation génétique de l'aire native et (iii) l'échantillonnage incomplet avec des zones envahies absentes, comme la Chine et la Corée du Sud. Les résultats non concluants obtenus soulignent enfin les limites de ces méthodes (pas de

calcul de probabilités, manque d'évaluation de la confiance statistique dans les résultats tels que des calculs d'erreurs de type I et II) et le besoin, dans notre cas, de méthodes plus puissantes, comme l'ABC décrite dans la partie I.3.b.

❖ Au-delà de l'article...

Dans cette étude, nous avons obtenu des valeurs de F_{st} très élevées et nous avons supposé que la faible variabilité intra-échantillon avait artificiellement augmenté les valeurs. Néanmoins, la littérature sur le sujet (Alcala et al. 2014; Balloux et al. 2000; Gregorious 2010; Hedrick 1999; Jost 2008; Jost 2009; Verity & Nichols 2014), l'examen plus précise de nos données et des essais de mesure d'autres indices de différenciation tel que le D de Jost (2008), nous a conduit à réaliser que la faible diversité intra-échantillon (et notamment la présence de nombreux allèles fixés) pouvait réellement perturber le F_{st} , au même titre qu'une forte diversité intra-échantillon. Par comparaison avec le D , nous nous sommes rendu compte que l'augmentation des valeurs de F_{st} , due à la faible diversité intra-échantillon, pouvait être considérable. Nous avons ainsi constaté que selon la mesure de différenciation utilisée, nous obtenions des résultats différents. Par exemple, le F_{st} entre l'échantillon portugais "AM2" et l'échantillon japonais "Kasumig5" était de 0.99 alors que le D était de 0.033 pour cette même comparaison. Nous n'avons pas ajouté le D dans cet article au moment de la révision, parce que cela ne changeait pas la discussion et les conclusions générales. En revanche, les deux mesures de différenciation (F_{st} et D) ont été calculées systématiquement par la suite dans la mesure où elles ne conduisent pas toujours aux mêmes conclusions et ne renseignent apparemment pas sur les mêmes processus (Alcala et al. 2014; Gerlach et al. 2010; Jost 2009).

Les différences qu'il peut exister entre les mesures de différenciation sont présentées très brièvement ici pour faire part de nos observations mais ce sujet est abordé plus longuement dans la discussion générale, (partie IV.2.a).

Article n°2

Mallez S, Castagnone C, Espada M, Vieira P, Eisenback J D, Harrell M, Mota M, Aikawa T, Akiba M, Kosaka H, Castagnone-Sereno P, Guillemaud T (2014). Worldwide invasion routes of the pinewood nematode: what can we infer from population genetics analyses? *Biological Invasions*. doi: 10.1007/s10530-014-0788-9 (online first)

Worldwide invasion routes of the pinewood nematode: What can we infer from population genetics analyses?

Sophie Mallez · Chantal Castagnone · Margarida Espada · Paulo Vieira · Jonathan D. Eisenback · Mark Harrell · Manuel Mota · Takuya Aikawa · Mitsuteru Akiba · Hajime Kosaka · Philippe Castagnone-Sereno · Thomas Guillemaud

Received: 27 March 2014 / Accepted: 22 September 2014
© Springer International Publishing Switzerland 2014

Abstract Identifying the invasion routes and determining the origin of new outbreaks of invasive species are of crucial importance if we are to understand the invasion process, improve or establish regulatory measures and, potentially, limit the damage. We focused here on the invasion of Europe by the pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle 1970; Nematoda: Aphelenchoididae), a major pest of forest ecosystems, native to North America and already invasive in Asia since the beginning of the twentieth century. We evaluated the genetic diversity and structure of worldwide field PWN samples by classical and Bayesian population genetics methods to deter-

mine the source of the European invasive populations and the number of introduction events in Europe. We found (1) a very strong spatial genetic structure in native PWN populations, (2) a very low level of polymorphism in each of the invaded areas and (3) contrasted results concerning the origin of European invasive populations. Our findings provide evidence for: (1) a large effect of genetic drift on the biological cycle of the PWN, due to intense demographic bottlenecks during tree infections, not compensated for by effective dispersal of its vector; (2) a single introduction event for each of the invaded areas in Japan and Europe and a small effective size for the introduced populations and (3) a mainland Portuguese origin for PWN populations from Madeira. However, more sophisticated methods of invasion route inference and broader sampling are required to conclusively determine the origin of the European outbreak.

Electronic supplementary material The online version of this article (doi:10.1007/s10530-014-0788-9) contains supplementary material, which is available to authorized users.

S. Mallez (✉) · C. Castagnone · P. Castagnone-Sereno · T. Guillemaud
UMR 1355 Institut Sophia Agrobiotech, INRA,
06903 Sophia Antipolis, France
e-mail: sophie.mallez@sophia.inra.fr

S. Mallez · C. Castagnone · P. Castagnone-Sereno · T. Guillemaud
UMR Institut Sophia Agrobiotech, Université de Nice
Sophia Antipolis, 06903 Sophia Antipolis Cedex, France

S. Mallez · C. Castagnone · P. Castagnone-Sereno · T. Guillemaud
UMR 7254 Institut Sophia Agrobiotech, CNRS,
06903 Sophia Antipolis Cedex, France

M. Espada · P. Vieira · M. Mota
NemaLab/ICAAM – Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora,
Núcleo da Mitra, Ap. 94, 7002-554 Évora, Portugal

J. D. Eisenback
Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA 24061, USA

M. Harrell
Nebraska Forest Service, University of Nebraska, Lincoln, NE 68583-0815, USA



Keywords *Bursaphelenchus xylophilus* · Population genetics · Diversity · Genetic structure · Microsatellite markers · Pine wilt disease

Introduction

Several studies have shown that it is more effective to fight biological invasions in their initial stages than to try to eradicate the invader after its establishment, in terms of cost, time and management efficiency (Allendorf and Lundquist 2003; Simberloff et al. 2013). Deciphering geographic pathways followed by invasive species is a key first step in this direction. Many studies carried out over the last decade have tried to identify the invasion routes of several invasive species, including insects (Papura et al. 2012; Pascual et al. 2007; Perdereau et al. 2013; Zepeda-Paulo et al. 2010), plants (Kelager et al. 2013), nematodes (Boucher et al. 2013) and fungi (Fontaine et al. 2013). Indeed, this approach can be used to address various questions of both practical and theoretical concern (Estoup and Guillemaud 2010). Specifically, from a practical point of view, it may facilitate the design of strategies for preventing new invasions, by highlighting weaknesses in control and phytosanitary measures. It may help to improve the control or eradication of detrimental invasive species, through the identification of natural enemies in the native area, better understanding of the biology of populations and definition of the ecological features of invasive populations (Mack et al. 2000; Strong and Pemberton 2000; Tsutsui et al. 2000; van Wilgen et al. 2013). From a more fundamental point of view, the inference of invasion routes may provide useful information about the invasion process. The testing of ecological

or evolutionary hypotheses explaining the success of biological invasions requires relevant comparisons between native and invasive populations (Facon et al. 2006; Keller and Taylor 2008; Puth and Post 2005; Wilson et al. 2009) and, thus, basic knowledge of the invasion routes used.

Here, we focused on the invasions of Europe by the pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner and Buhner 1934; Nickle, 1970; Nematoda: Aphelenchoididae), a microscopic worm that reproduces sexually (Futai 2013). Outside North America (widely recognized as the native range of this species, Dropkin et al. 1981; Kiritani and Morimoto 2004; Wingfield et al. 1982), PWN is the causal agent of pine wilt disease (Mamiya 1972, 1976, 1983) and it poses a serious threat to pine forests worldwide, due to ecological and economic consequences of infestation (Mamiya 1988; Soliman et al. 2012; Suzuki 2002; Vicente et al. 2011). It was first observed outside its native range in Japan, in 1905, near Nagasaki (Mamiya 1988), where 28 % of the 2.1 million ha of pine forest was found to be infested in 2000 (Mamiya 2004). It has since spread to other Asian countries, including China, Taiwan and South Korea (Moon et al. 2007). PWN was detected in the Setúbal Peninsula, close to Lisbon in Portugal, in 1999 (Mota et al. 1999) and new outbreaks have been identified since 2008 in the centre of mainland Portugal and on Madeira (Fonseca et al. 2012), as well as in Spain (Abelleira et al. 2011; Robertson et al. 2011). PWN was designated a quarantine organism by the European and Mediterranean Plant Protection Organization (OEPP/EPPO 1997) and the European Commission ruling (2006/133/CE) has imposed strict measures on the trade of wood, to limit the invasion. These measures have added significant costs to those already resulting from the destruction of pine forests.

Many studies have attempted to establish the origin of invasive outbreaks of PWN in different geographic areas (Cheng et al. 2008; Figueiredo et al. 2013; Fonseca et al. 2012; Metge and Burgermeister 2008; Pereira et al. 2013; Tares et al. 1992; Valadas et al. 2012a, b; Vieira et al. 2007; Zhang et al. 2008; Zhou et al. 2007). An Asian origin for European invasive populations of PWN was thus proposed (Figueiredo et al. 2013; Fonseca et al. 2012; Metge and Burgermeister 2008; Valadas et al. 2012b). However, these studies were subject to experimental limitations, such as the use of low-resolution markers (e.g. RAPD,

A27 T. Aikawa
A28 Tohoku Research Center, Forestry and Forest Products
A29 Research Institute, Morioka, Iwate 020-0123, Japan

A30 M. Akiba
A31 Forest Pathology Laboratory, Forestry and Forest
A32 Products Research Institute, Matsunosato 1, Tsukuba,
A33 Ibaraki 305-8687, Japan

A34 H. Kosaka
A35 Kyushu Research Center, Forestry and Forest Products
A36 Research Institute, Kurokami 4-11-16,
A37 Kumamoto 860-0862, Japan

RFLP), pools of individuals and/or collection samples. Thus, no study has as yet clearly investigated the worldwide invasion routes of PWN with relevant molecular markers, appropriate samples and the correct population genetics approaches as described by Estoup and Guillemaud (2010).

Therefore, the aim of this study was to perform population genetics analyses on natural PWN populations from around the world, with microsatellite markers developed previously (Mallez et al. 2013). In particular, we aimed to confirm the existence of the strong genetic structure in the native area of this species suggested by Mallez et al. (2013), through the use of a larger number of samples from different US states and hierarchical sampling. We also evaluated the genetic diversity of PWN populations from invaded areas, to obtain information about the number of introduction events. Finally, we investigated the relationships between populations in native and invaded areas, to clarify the invasion routes used by PWN. In particular, we investigated whether the European invasion resulted from an introduction independent of the Asian introduction or from successive introductions via Asia. We also analysed the relationships between invasive populations from mainland Portugal and Madeira.

Materials and methods

The biological system of the PWN and its generation time

The biological cycle of the PWN in natura is complex and involves at least two partners: the pine tree and the insect vector (Evans et al. 1996; Futai 2013 for a review). The PWN and its vector are closely associated and their life cycles closely match (Mamiya 1972). Thus, the PWN reproduce exclusively during summer once it invaded a susceptible pine tree (from June to September for the largest period). Its generation time is not precisely known in natura due to impossible direct observations of the PWN into the tree. However, from laboratory experiments, it was determined that the life cycle of the PWN depends mainly on the temperature, which determines its duration (Mamiya 1975). The generation time of the PWN may thus be very short and lasts from 4–5 days at 25 °C to about 12 days at 15 °C, with temperature

thresholds for development of 9.5 °C for the minimum and 33 °C for the maximum (Mamiya 1975). Consequently, by coupling this information, the generation time of the PWN was approximated to 30 generations per year (by averaging the summer temperature to 25 °C and by considering the June–September period for PWN reproduction). This is about 450 generations since the PWN introduction in Portugal and more than 3,200 generations since its introduction in Japan.

Sampling

Nematode samples were obtained from three different geographic areas: a part of the native area, the USA and two invaded areas, Japan and Portugal/Madeira. Thirty-four locations were sampled and 770 individuals were analysed in total: 18 locations from the USA (391 individuals), nine from mainland Portugal and Madeira (169 individuals) and seven from Japan (210 individuals). The PWN samples are listed and described in Table 1 and the locations from which the samples were collected are shown in Fig. 1. All individuals were extracted from wood samples collected directly from field locations. Each location sample originated from a single tree and consisted of seven to 36 individuals, at various stages. Nematodes were extracted with a sieve or a Baermann funnel (Viglierchio and Schmitt 1983). No permission was required to collect samples of PWN from the infested areas and we obtained an official agreement from the French authorities (#2012060-0004) for the importation and manipulation of this quarantine organism at Institut Sophia Agrobiotech facilities. For some samples, extraction of individuals were carried out by our collaborators, who then sent them in DESS (Yoder et al. 2006). These samples were washed in distilled water before DNA extraction. For the other samples, DNA was extracted directly after the extraction of individuals from wood samples.

DNA extraction and genotyping of microsatellite loci

Single individuals were subjected to a thermal shock, as explained below, for DNA extraction (Castagnone et al. 2005). Each individual was hand-picked, transferred to 18 µl of lysis buffer (10× *Taq* buffer with MgCl₂, *Taq* Core Kits10, MP Biomedicals; 60 mg ml⁻¹ proteinase K and sterile distilled H₂O) and placed at -80 °C for

Table 1 Characteristics and population genetics summary statistics of each sample of *B. xylophilus* used in this study

Sample code	No. of individuals	Origin	Mean Na	Max Na	He	Ho	Fis
MO1	31	USA—Missouri—Columbia	3.25	8	0.33	0.27	0.19 ^a
MO2	23	USA—Missouri—Columbia	3.38	7	0.36	0.26	0.27 ^a
NE1	16	USA—Nebraska—Davey	2.00	4	0.20	0.18	0.09
NE2	15	USA—Nebraska—Davey	1.81	4	0.21	0.16	0.24
NE5	14	USA—Nebraska—Pawnee Lake	1.94	5	0.21	0.18	0.14
NE6	21	USA—Nebraska—Pawnee Lake	1.38	2	0.18	0.10	0.48 ^a
NE9	29	USA—Nebraska—Pawnee Lake	2.63	5	0.31	0.23	0.26 ^a
NE10	26	USA—Nebraska—Pawnee Lake	1.75	4	0.19	0.15	0.19 ^a
NE12	28	USA—Nebraska—Conestoga Lake	1.50	3	0.15	0.10	0.34 ^a
NE13b	19	USA—Nebraska—Pioneers Park	2.38	5	0.28	0.20	0.31 ^a
NE14	28	USA—Nebraska—Pioneers Park	1.69	3	0.24	0.16	0.34 ^a
NE15	23	USA—Nebraska—Pioneers Park	2.25	5	0.22	0.14	0.38 ^a
NE19	16	USA—Nebraska—UNL East Campus	1.81	3	0.23	0.20	0.15
NE22	17	USA—Nebraska—Lincoln	2.00	5	0.28	0.20	0.28 ^a
NE23	25	USA—Nebraska—Lincoln	1.25	2	0.07	0.07	0.09
NE24	19	USA—Nebraska—Lincoln	1.31	2	0.07	0.06	0.15
VI9	22	USA—Virginia—Midlothian	1.50	3	0.16	0.14	0.14
MA1	19	USA—Massachusetts—Worcester	1.50	2	0.14	0.14	0.00
Jap120	23	Japan—Iwate—Shiwa	1	1	—	—	—
Jap212	27	Japan—Iwate—Shiwa	1	1	—	—	—
Jap308	25	Japan—Iwate—Shiwa	1	1	—	—	—
Kasumig2	36	Japan—Ibaraki—Kasumigaura	1	1	—	—	—
Kasumig3	29	Japan—Ibaraki—Kasumigaura	1.19	2	0.06	0.06	−0.02
Kasumig5	35	Japan—Ibaraki—Kasumigaura	1	1	—	—	—
Kosa	35	Japan—Kumamoto—Kosa	1.25	2	0.06	0.05	0.03
Mad23PC	12	Madeira Island—Porto da Cruz	1	1	—	—	—
Mad24C	7	Madeira Island—Calheta	1	1	—	—	—
128S	17	Portugal—Setubal—Grândola	1.06	2	0.03	0.01	0.62 ^a
TR1	30	Portugal—Setubal—Troia	1	1	—	—	—
TR2	27	Portugal—Setubal—Troia	1	1	—	—	—
AM2	21	Portugal—Setubal—Aguas de Moura	1	1	—	—	—
Comporta	28	Portugal—Setubal—Comporta	1	1	—	—	—
E182	13	Portugal—Coimbra—Penela	1	1	—	—	—
E1069	14	Portugal—Viseu—Castro Daire	1	1	—	—	—

Mean Na is the mean number of alleles per sample over all loci, Max Na is the maximum number of alleles per locus in each sample, He is the expected heterozygosity and Ho is the observed heterozygosity. *FIS* was calculated as described by Weir and Cockerham (1984)

“—” indicates that Ho, He and *FIS* were not calculated, for samples with only monomorphic markers

^a That the result of the HWE test was significant at the 5 % level after FDR correction (Benjamini and Hochberg 1995)

214 60 min. It was then immediately transferred to 60 °C for
 215 60 min and, finally, to 95 °C for 15 min in a Biometra®
 216 T3-Thermoblock Thermocycler. We amplified 16
 217 microsatellite loci in three multiplex PCRs: MA28 (5

microsatellite loci), MB28 (5 microsatellite loci) and
 MC33 (6 microsatellite loci), as described by Mallez
 et al. (2013). We excluded Bx07 from the MB28
 multiplex reaction, because this marker is identical to

218
 219
 220
 221

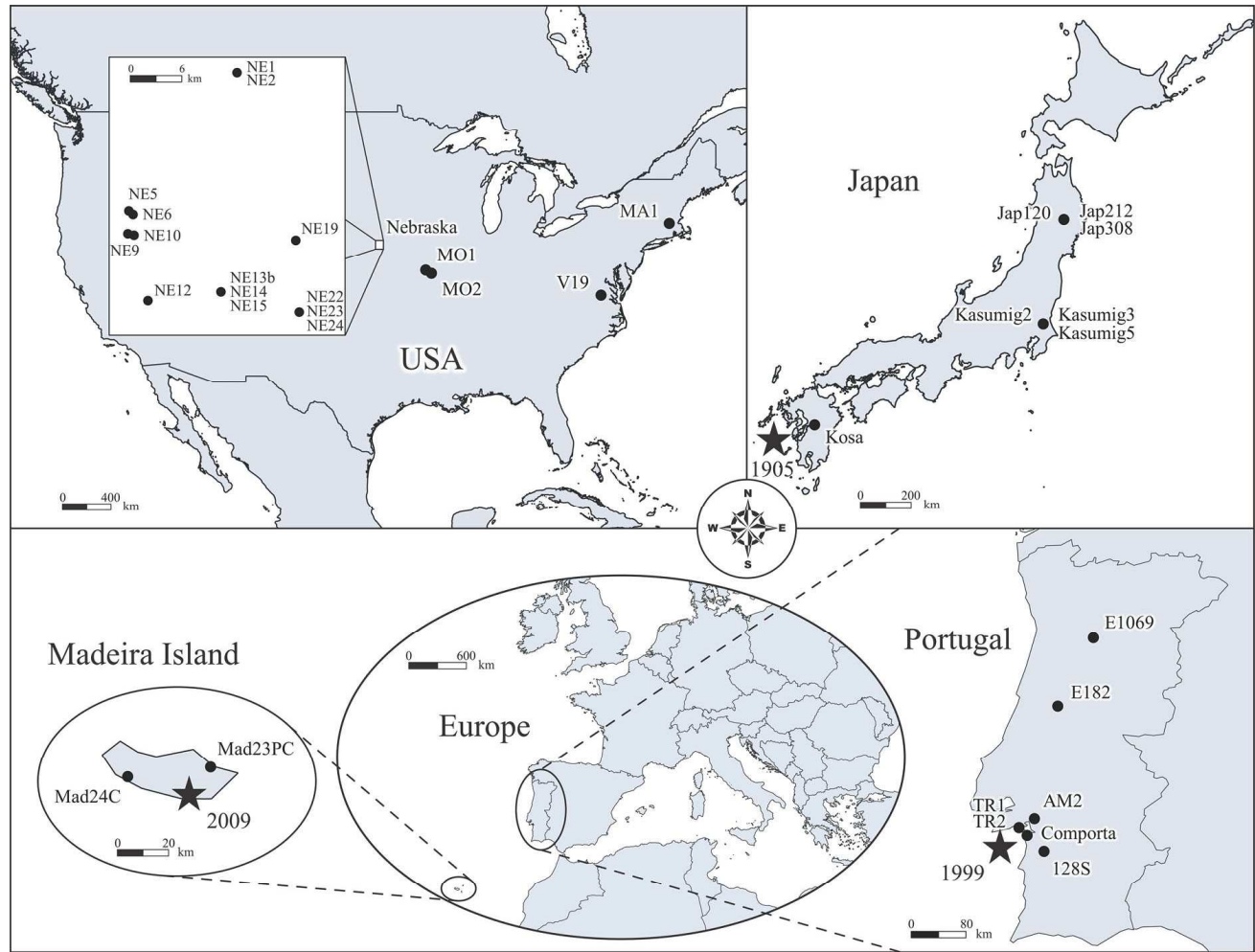


Fig. 1 Location of the sampling sites for *B. xylophilus* used in this study. Three different geographic areas are represented: the USA, a part of the native area, focusing on the Nebraska, and two different invaded areas, mainland Portugal and Madeira and

Japan. The stars indicate the hypothetical entry points of PWN. The dates of first observation for each invaded area are indicated. The codes on the maps are the sample names. For more details, see Table 1

PWN_51 (Mallez et al. 2013). PCR amplifications were performed in 10 μ l containing 1 \times QIAGEN Multiplex Master Mix, 2 μ M of each primer, with forward primers labelled with a fluorescent dye (6-FAM, VIC, PET or NED) at the 5' end, and 2 μ l of genomic DNA extracted by thermal shock. The amplification reactions were performed in a Biometra[®] T3-Thermoblock Thermocycler and included a 15 min denaturation step at 95 $^{\circ}$ C, followed by 28 or 33 cycles (depending on the multiplex PCR) of 30 s at 94 $^{\circ}$ C, 1.5 min at 55 $^{\circ}$ C, and 1 min at 72 $^{\circ}$ C, followed by a final extension step of 30 min at 60 $^{\circ}$ C. Genotypes were determined with an ABI 3700 sequencer (Applied Biosystems), with the 500 LIZTM GeneScanTM size standard (Applied Biosystems) and GenemarkerTM version 1.75 software (SoftGenetics LLC).

Data analyses 238

Standard genetic analyses 239

For each sample, we determined the maximum number of alleles detected over all loci (Max. Na), the mean number of alleles (Mean Na) and the observed (H_o) and expected (H_e) heterozygosities per sample, with GENETIX version 4.05 (Belkhir et al. 1996–2004). We evaluated deviation from Hardy–Weinberg equilibrium (HWE) with GENEPOP version 4.1.3 (Rousset 2008) and we quantified inferred deviations from HWE by calculating the Weir and Cockerham estimate of F_{IS} (Weir and Cockerham 1984) with FSTAT version 2.9.3.2 (Goudet 2002). Linkage equilibrium between loci was assessed with 251

the log likelihood ratio test, in GENEPOP (Rousset 2008). We took multiple testing (HWE tests) and the non-independence of tests (linkage tests) into account by performing false discovery rate (FDR) correction (Benjamini and Hochberg 1995) and sequential Bonferroni adjustment (Sokal and Rohlf 1995), respectively. Null allele frequencies were estimated with FREENA (Chapuis and Estoup 2007).

Genetic structure analyses within areas

All the analyses described in this section were performed within each area, for precise characterization of the nematode samples from the native area and the two invaded areas. We first tested the hypothesis of uniform genotype frequencies between samples, by Fisher's exact test (Raymond and Rousset 1995) implemented in GENEPOP (Rousset 2008). As non-independent multiple tests were performed, sequential Bonferroni correction (Sokal and Rohlf 1995) was carried out to adjust significance levels. We also calculated Weir and Cockerham estimates of F_{ST} (Weir and Cockerham 1984) between samples, corrected for null alleles with FREENA (Chapuis and Estoup 2007). We then studied the structure of the populations from each area by a Bayesian clustering approach, as implemented in STRUCTURE version 2.3 (Pritchard et al. 2000). An admixture model with correlated allele frequencies was used (Falush et al. 2003). The numbers of clusters tested, K , varied from 1 to 18 for the USA, from 1 to 7 for Japan and from 1 to 4 for Portugal/Madeira. We carried out 20 independent runs for each value of K . Each run involved a Markov Chain Monte Carlo (MCMC) procedure with 10^6 iterations, following a burn-in period of 2×10^5 iterations. Default values were maintained for all the other parameters. The number of clusters was determined both as described by Evanno et al. (2005) and automated in STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/struct_harvest/index.php, Earl and vonHoldt 2012) and by checking all the bar plots of co-ancestry parameters for successive values of K . Using CLUMPP (Jakobsson and Rosenberg 2007), we identified the most frequent clustering patterns for each value of K among the 20 runs, which we plotted with DISTRICT version 1.1 (Rosenberg 2004). We checked the adequacy of the clustering patterns chosen for successive values of K .

In the USA, the hierarchical sampling in Nebraska allowed us to investigate the possible occurrence of isolation by distance (IBD). This involved assessment of the correlation between genetic distance ($F_{ST}/(1 - F_{ST})$) and the logarithm of geographic distance, for pairs of populations (Rousset 1997). We used the Mantel test in GENEPOP (Rousset 2008) and 20,000 permutations to assess the significance of the correlation.

Relationships between the different areas

We focused on the origin of the Portuguese mainland and island (Madeira) samples. The most probable source population, in North America or in Asia, for each Portuguese sample was investigated in several ways. More precisely, the aim of this part was to select the most probable scenario among the following scenarios: (1) a scenario with two independent introductions, in Asia and in Europe, from North America and (2) a scenario with two successive introductions, from North America to Asia and then from Asia to Europe.

We first analysed the F_{ST} values corrected for null alleles (Chapuis and Estoup 2007; Weir and Cockerham 1984) between each Portuguese sample and each American or Japanese sample. We then carried out an individual assignment likelihood analysis (Paetkau et al. 2004; Rannala and Mountain 1997), as in previous studies on invasion routes (Ciosi et al. 2008; Pascual et al. 2007) with GENECLASS2 software version 2.0 (Piry et al. 2004). This analysis involves calculating the mean individual assignment likelihood (denoted $L_{i \rightarrow s}$) of each Portuguese sample i , to each possible source population s (the American samples and the Japanese samples, in our case). The most probable source of a target invasive population sample i is considered to be the population with the lowest corrected F_{ST} values with i and the maximum assignment likelihood of i . We expect a lower corrected F_{ST} value between the USA and Portugal/Madeira Island than between Asia and Portugal/Madeira Island under the scenario of independent events of introduction and the opposite under the scenario of successive events of introduction.

We also plotted a neighbour joining (NJ) tree (Saitou and Nei 1987), based on Cavalli-Sforza and Edwards genetic distances (Cavalli-Sforza and Edwards 1967) with POPULATION software version 1.2.30

(<http://bioinformatics.org/~tryphon/populations/>). The robustness of the tree topology was evaluated by carrying out 2,000 bootstrap replicates over loci. A tree based on Cavalli-Sforza and Edwards genetic distances (Cavalli-Sforza and Edwards 1967) corrected for null alleles (Chapuis and Estoup 2007) was also built. The most probable source of a target invasive population sample i is considered to be the population from outside Europe whose sample is clustering closest to i in the tree.

Finally, we carried out Bayesian clustering analysis with STRUCTURE software (Pritchard et al. 2000) using all the samples from North America, Asia and Europe to determine the origin of the sampled Portuguese populations. The number of clusters tested, K , varied from 1 to 10. As before, an admixture model with correlated allele frequencies (Falush et al. 2003), 20 runs per K , 10^6 iterations for the MCMC and 2×10^5 iterations for the burn-in period were used. CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004) was used to identify the most frequent clustering patterns for each value of K and to display the corresponding bar plots, respectively. This method identifies the most probable source of a target invasive population sample i as the population for which the sample(s) is (are) the last to cluster with i with increasing values of K . Thus, under the scenario of independent events of introduction, we expect that an American sample is the last one to still cluster with Portugal/Madeira Island with increasing values of K . Conversely, under the scenario of successive events of introduction, we expect an Asian sample to be the last one to still cluster with Portugal/Madeira Island with increasing values of K .

Results

Standard genetic analyses

Three markers were monomorphic (M3, M30 and M49), even in the native area (but see Mallez et al. 2013, who found polymorphism at these markers in the native area, in analyses of collection and field samples). Genetic diversity was low to moderate in the USA and low to extremely low in invaded areas (see Table 1). In the native area, we detected up to eight alleles per locus and per sample (and up to 13 alleles per locus over all samples), with a mean number of alleles per sample (Mean Na) of 1.31–3.38 and a mean

expected heterozygosity (H_e) of 0.07–0.36. Numerous monomorphic markers were detected in the invaded areas: 10 markers in Japan and 15 markers in Portugal/Madeira, of the 16 considered. No more than two alleles per locus and per sample (and no more than three per locus over all samples in Japan) were detected, with a mean Na of 1.25 and a mean H_e of 0.06 at most. Ten samples in the native area and one sample in Portugal deviated significantly from HWE because of a heterozygous deficit (see Table 1). Significant linkage disequilibrium was found in 10 of the 624 pairwise tests carried out (after sequential Bonferroni correction, Sokal and Rohlf 1995), for four pairs of loci: M62 and Bx08; M51 and M56; M35 and M56 and M35 and M51. Examination of the results of FREENA analysis (Chapuis and Estoup 2007) showed that most of the deviation from HWE (especially in Portugal) and most of the significant linkage tests were accounted for by the presence of null alleles at the loci involved (estimates of null allele frequencies from 7 to 24 %, data not shown). However, null alleles were not systematically observed for a given marker across all the samples or for a given sample across all the markers. In addition, none of the samples or markers had more than 10 % null alleles on average. We therefore used the entire dataset for further analyses.

Genetic structure analyses within geographic areas

In the USA, all samples displayed significant differentiation after correction for multiple testing (Fisher's exact tests, $p < 10^{-5}$). Pairwise corrected estimates of F_{ST} values were also very high: from 0.06 to 0.76 (for more details see Table S1). The number of clusters could not be clearly inferred from the Bayesian clustering analysis, because ΔK (Evanno et al. 2005) had several peaks, at different values of K (see Figure S1). A biologically meaningful genetic structure occurred for large values of K (see the examples of bar plots for several values of K in Fig. 2a and all bar plots in Figure S2). Most samples were progressively unambiguously assigned to different clusters, at least until $K = 14$. We also detected evidence of IBD (slope = 0.064, $p = 0.013$).

In Japan, the three samples from Iwate in the North were genetically identical, presenting one fixed allele for all the markers. All the remaining samples appeared to display significant differentiation (Fisher's exact tests, $p < 10^{-5}$), with extremely high

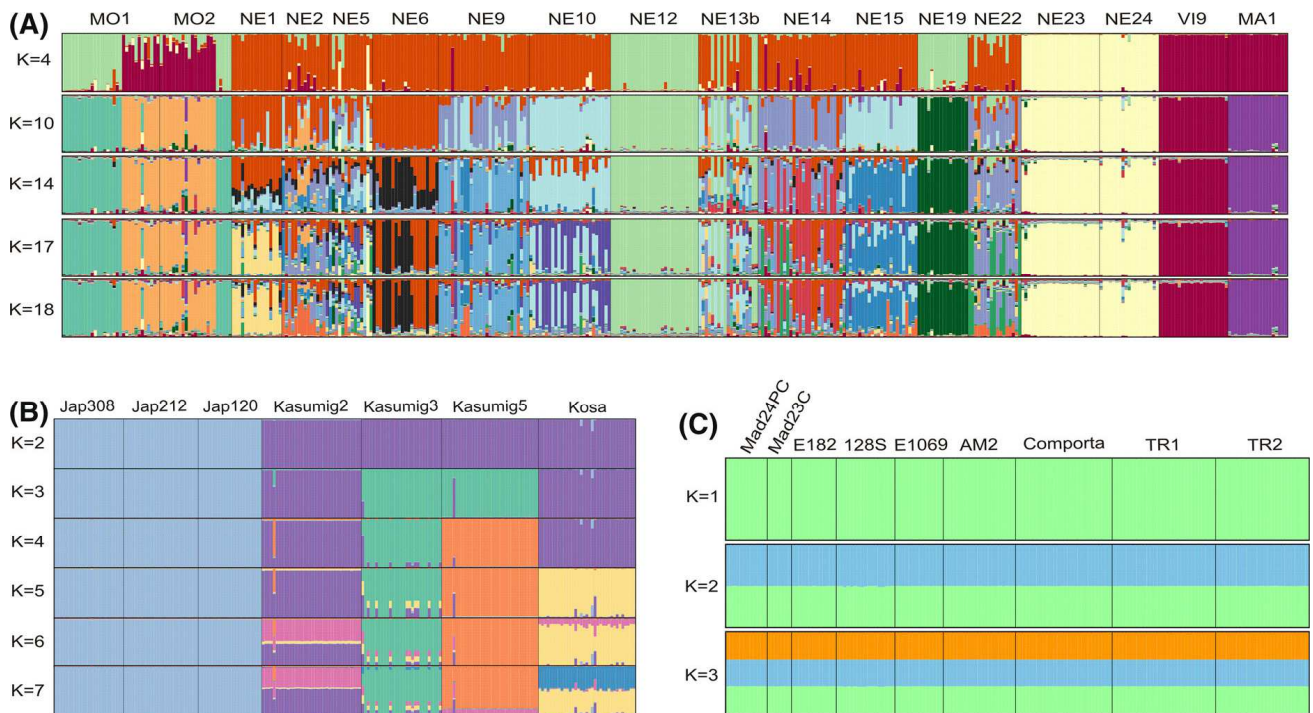


Fig. 2 Genetic structure of the PWN field samples within each area. Bar plots of the coefficients of co-ancestry obtained in various STRUCTURE analyses with several values of K for **a** a part of the native area, the USA and for the two invaded areas

studied here, **b** Japan and **c** Portugal/Madeira. Each bar corresponds to one individual nematode and each cluster is represented with a particular colour

corrected F_{ST} values, ranging from 0.63 to 0.99 (excluding the three identical samples, see Table S1 for more details). Strong genetic structure was also detected at the scale of the tree, with STRUCTURE. The ΔK method (Evanno et al. 2005) inferred the existence of three clusters (see Figure S1), but likelihood values reached a plateau at $K = 5$ and an examination of successive bar plots for co-ancestry suggested a meaningful structure for $K = 5$ (see Fig. 2b).

In Portugal, eight of the nine samples were genetically identical and 162 individuals (of the 169 sampled) had identical homozygous multi-locus genotypes. The remaining sample (128S) differed significantly from seven of the other eight samples after correction for multiple testing (Fisher's exact tests, $p < 0.017$). The result of one test was not significant (Mad24C vs. 128S, Fisher's exact test, $p = 0.065$), probably because of the small size (seven individuals) of the Mad24C sample. The clustering analysis inferred a single cluster, grouping together the samples from mainland Portugal and Madeira (see Fig. 2C, Figure S1).

Relationships between populations from the different geographic areas

The results of the various analyses performed to clarify the relationships between the populations in different geographic areas are visualized and summarized in Fig. 3. The lowest F_{ST} values obtained with Portuguese samples always corresponded to American samples (see the example of one Portuguese sample in Fig. 3a and all Portuguese samples in Figure S3). The mean F_{ST} value (across samples) between Portugal/Madeira and the USA was also lower than that between Portugal/Madeira and Japan, as shown by the dashed lines in Fig. 3a. Thus, the populations of Portugal/Madeira seem to be closer to the American populations than to the Japanese populations, on the basis of F_{ST} . Portuguese samples were assigned to American samples with the largest mean individual likelihood ($L_{i \rightarrow s}$, see the example of one Portuguese sample in Fig. 3b and all Portuguese samples in Figure S4). However, if we averaged across samples, Portugal was assigned to Japan with the largest mean individual likelihood (dashed lines on Fig. 3b). The

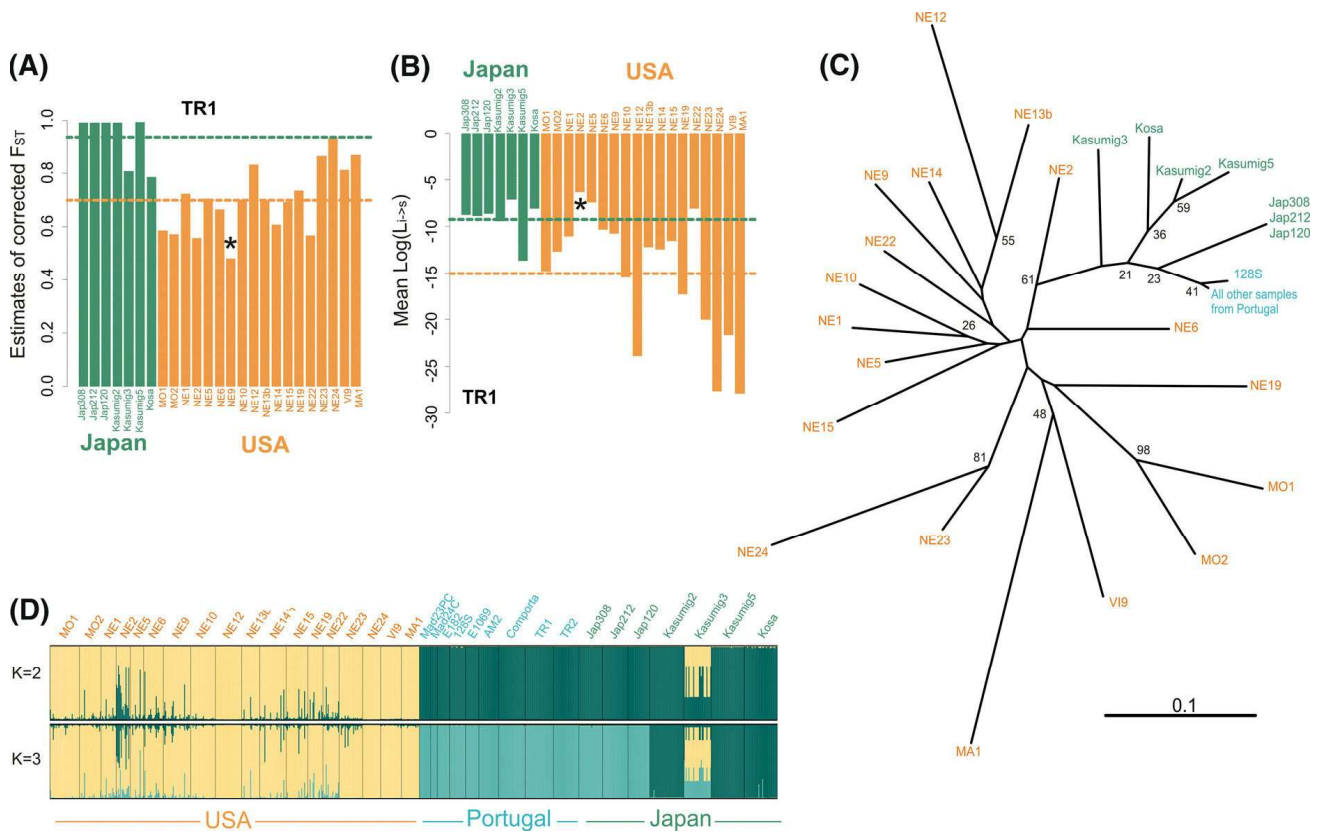


Fig. 3 Genetic relationships between populations from the different areas. The results of the four analyses performed are shown. **a** Weir and Cockerham estimates of F_{ST} corrected for null alleles (Chapuis and Estoup 2007; Weir and Cockerham 1984) between TR1, a Portuguese sample shown as an example, and each Japanese (in green) or American (in orange) sample. **b** Mean log-likelihood of the multilocus individual assignment ($L_{i \rightarrow s}$) of TR1 to each Japanese (in green) or American (in orange) sample. For these first two analyses, the sample displaying the lowest F_{ST} with TR1 or the largest mean $L_{i \rightarrow s}$ is indicated with an asterisk. The mean values across samples

for each area and for each parameter are represented by dashed lines. The results of F_{ST} and $L_{i \rightarrow s}$ for other Portuguese samples are given in Figure S1 and Figure S2, respectively. **c** NJ tree based on Cavalli-Sforza and Edwards distances (Cavalli-Sforza and Edwards 1967). Bootstrap (on locus) values calculated over 2,000 replications are given as percentages (only values >20 % are shown). American samples are shown in orange, Japanese samples in green and Portuguese samples in blue. **d** Bar plots of the coefficients of co-ancestry obtained with STRUCTURE for the first values of K . Each bar corresponds to one individual nematode and each cluster is represented by a particular colour

origin of the Portuguese samples is thus not clearly inferred with this statistic. On the NJ tree, the Portuguese samples were closer to the Japanese samples than to the American samples, for trees generated with both uncorrected and corrected Cavalli-Sforza and Edwards distances (Cavalli-Sforza and Edwards 1967; Chapuis and Estoup 2007; Fig. 3c, Figure S5, respectively). This result was robust to the use of other genetic distances (data not shown). Finally, Bayesian clustering analysis (Fig. 3d) showed that the Portuguese samples clustered with the Japanese samples at the lowest level of structuration ($K = 2$), with the Portuguese samples and three Japanese samples subsequently clustering together,

separately from the other Japanese samples, at $K = 3$. This result was confirmed by a clustering analysis of the samples from invaded areas only (Portugal/Madeira and Japan, data not shown).

Discussion

In this study, we investigated the genetic diversity and structure of natural PWN populations from around the world and the relationships between populations from native and invaded areas. Three main results were obtained: (1) we confirmed here, with a much larger number of samples, the existence of the strong genetic

structure of PWN populations suggested by Mallez et al. (2013), consistent with notion that genetic drift has had a major effect on the genetic structure of PWN, (2) we found very low levels of polymorphism in the invaded areas, suggesting single introduction events, introduced populations with small effective sizes and clarifying the relationships between invasive populations within Europe and (3) we observed that classical and Bayesian population genetics methods were inconclusive concerning the invasion routes followed by PWN, and that more powerful inference methods are therefore required.

Strong genetic structure in PWN populations

We observed a strong spatial genetic structure of native PWN populations. A very low genetic diversity within samples from the native area compared to that between populations, highly significant differentiation tests, Bayesian clustering analysis and the extremely high values of pairwise F_{ST} were indicative of this structure. The very low within sample variability may have artificially increased F_{ST} values.

Such a genetic structure has already been reported in the native ranges of invasive species (e.g. for a Cuban lizard, by Kolbe et al. 2004). Our findings confirm those reported by Mallez et al. (2013), for a larger number of samples with hierarchical sampling in Nebraska. They confirm that the migration-drift equilibrium is highly biased towards significant genetic drift without compensation by efficient dispersal, even over short distances. The limited dispersal of the PWN may reflect the complexity of its dispersal process, which is principally dependent on an insect vector, *Monochamus* species (Akbulut and Stamps 2011; Linit 1988; Mamiya and Enda 1972; Sousa et al. 2001). This insect probably has a weak dispersal capacity, resulting in dispersal over short distances (Shibata 1986; Togashi 1990), particularly when the vector is heavily loaded with nematodes, potentially affecting its ability to fly (Akbulut and Linit 1999). The evidence of genetic IBD is also consistent with the probably spatially limited dispersal capacity of the PWN.

PWN must also overcome several potential obstacles to its dispersal. It must aggregate around pupal chambers, entering the beetles just before emergence, facilitating its entry into the tree via the maturation feeding sites of the beetles, for effective reproduction

within the tree (reviewed in Futai 2013; Vicente et al. 2011). These steps may also contribute to the strong genetic drift observed in this system, because (1) the population of nematodes in an infested tree is aggregated into a small number of beetles that emerge from the tree (Akbulut and Linit 1999), (2) most beetles carry only a bit of nematodes (Kobayashi et al. 1984; Linit 1988) and (3) not all nematodes successfully invade pine trees (Togashi 1985), resulting in relatively low transmission rates (10–20 %, Kobayashi et al. 1984; Togashi 1985).

Thus, the initial PWN population in a tree may be small, subsequently increasing exponentially in size. This creates considerable genetic drift, coupled with rates of dispersal between trees too low to homogenise genetic diversity.

Very low levels of polymorphism in invaded areas

Another finding in this study was the very low level or even complete absence of genetic diversity in the invaded areas, as observed in previous reports on the same species (Fonseca et al. 2012; Pereira et al. 2013; Vieira et al. 2007; Zhang et al. 2008; Zhou et al. 2007). However, this finding is remarkable in comparison with other invading species in which a loss of genetic diversity during invasion is not as common as previously expected (Bossdorf et al. 2005 for a review in plant invasive species; Roman and Darling 2007 for a review in aquatic invasions). The large number of monomorphic markers and the similarity between them across individuals (fixed alleles) are surprising at first glance for a sexual species (Futai 2013) and for microsatellite markers. However, these findings may be accounted for by the low level of intra-sample genetic diversity observed in the native area, together with the genetic bottlenecks and founder events often occurring during the introduction of species in new areas (Allendorf and Lundquist 2003; Sakai et al. 2001). Moreover, biological invasions tend to occur over short timescales, so mutational processes have very little effect on the genetic structure of invasive populations in the short term. Consequently, genetic structure is shaped mostly by demographic processes, such as intense demographic bottlenecks resulting in intense genetic bottlenecks, as in this study.

This very low level of genetic diversity provides information about the most probable number of introduction events. In cases of multiple introductions

in a restricted area, such as mainland Portugal, Madeira or Japan, significant genetic diversity would be expected, due to admixture of the various introduced populations, particularly given the highly structured nature of native populations. The almost complete absence of polymorphism in mainland Portugal, Madeira and Japan therefore strongly suggests that a single introduction occurred in each of these areas, with each introduced population having small effective size. In the European context, these findings firmly suggest that the second outbreak detected in the centre of mainland Portugal in 2008 resulted from expansion of the first outbreak detected close to Lisbon in 1999. This conclusion contrasts with the findings of Valadas et al. (2012a), who suggested that multiple introductions had occurred. Our findings also suggest that the PWN populations on Madeira originated from mainland Portugal, given the near identity of the populations from Madeira and mainland Portugal and the first detection of PWN outbreak on Madeira 10 years after the first outbreak in mainland Portugal.

Intra-population genetic diversity is widely considered to be a prerequisite for adaptation to changing conditions and/or environment (Reed and Frankham 2003; Willi et al. 2006). Biological invasions have thus brought to light a genetic paradox (Allendorf and Lundquist 2003; Frankham 2005): the occurrence of successful invasions with low levels of genetic diversity. The case studied here provides a good example of this paradox, because PWN populations with low levels of diversity have managed to invade several regions around the world. There are several ecological mechanisms that might account for this paradox in the case of PWN: (1) the presence of appropriate insect vectors in each of the countries invaded (Akbulut and Stamps 2011; Mamiya and Enda 1972; Naves et al. 2007; Sousa et al. 2001), (2) the presence of susceptible hosts (Evans et al. 1996) and (3) the greater competitiveness of PWN than of its closely relative resident in the area, *B. mucronatus* (Cheng et al. 2009; Vincent et al. 2008). There is also some published evidence that a loss of genetic diversity (demonstrated with neutral markers) does not hamper the adaptive phenotypic variation of fitness-related traits (Dlugosch and Parker 2008) and that measurements of neutral genetic diversity are of only limited value for the prediction of quantitative genetic variability (Reed and Frankham 2001). Further

studies are therefore required to determine whether invading PWN populations display significant adaptive genetic variability and whether this variability contributes to the success of PWN.

Worldwide invasion routes of PWN: the need for more powerful methods

The third main finding of this study was the difficulty elucidating the worldwide invasion routes of PWN. Depending on the analysis and the method used, we alternatively inferred two possible origins, North America and Asia, for the Portuguese outbreaks. The *FST* and mean *FST* values across samples suggested an American origin for all the Portuguese samples, whereas the NJ tree and the Bayesian clustering analysis suggested a Japanese origin for these samples. A discrepancy was also found in the mean individual assignment likelihood analysis, which gave inconsistent results, as the sample with the minimum mean individual assignment likelihood suggested an American origin and the mean individual assignment likelihood between samples suggested an Asian origin. This discrepancy partly results from large inter-sample variance of allelic frequency in North America. Previous studies have proposed an Asian origin for the European populations of PWN (Fonseca et al. 2012; Metge and Burgermeister 2008; Valadas et al. 2012b), essentially on the basis of tree analyses. Our finding with the NJ tree and Bayesian clustering is consistent with this conclusion. One key point here is that the conclusions drawn from these previous studies are no more robust than ours because (1) the methods used are included among those used here and (2) these previous studies did not use other methods possibly leading to alternative conclusions.

These inconclusive results highlight a major problem with traditional methods: a lack of statistical confidence evaluation for inferences of the source population of invasion. No statistical tests are carried out and no probabilities or type I or type II errors are calculated for classical and Bayesian clustering or distance methods. This makes it difficult to determine which result is the most likely when several alternatives are proposed (Estoup and Guillemaud 2010). However, these classical and Bayesian methods have proved both useful and conclusive in other cases of invasion (Ciosi et al. 2008; Facon et al. 2003; Kolbe et al. 2004; Papura et al. 2012; Perdereau et al. 2013; Wan et al. 2012). We can therefore put forward several

hypotheses to account for the conflicting results obtained here: (1) the low level of diversity may have resulted in a lack of power to discriminate between the two possible alternatives, (2) the strong spatial genetic structure observed in the USA, requiring a very large sampling scheme to embrace most genetic variation of the native and (3) the lack of samples from some of the existing invaded areas, such as China or South Korea, potentially serving as sources for the invasion of Europe, as suggested by Figueiredo et al. (2013) on the basis of analyses of collection samples for PWN.

Given the limitations of classical and Bayesian methods, the use of recent model-based methods, such as the approximate Bayesian computation (ABC, Beaumont et al. 2002; Bertorelle et al. 2010; Guillemaud et al. 2010) may prove useful. The enthusiasm linked to the use of genetic data for reconstructing the history of invasive species was restricted by failures or technical limitations (Barun et al. 2013; Fitzpatrick et al. 2012). However, ABC make possible to perform extensive simulations of various/alternative hypothesis, which is needed to make reliable biological interpretations of invasion (Barun et al. 2013). Moreover, ABC offers several advantages (described by Estoup and Guillemaud 2010) that may be crucial in studies of PWN: (1) it takes complex scenarii into account, (2) it manages incomplete sampling by providing the possibility of considering unsampled “ghost” populations and, most importantly, (3) it makes it possible to evaluate quantitatively and to compare statistically the various competing scenarii, through the calculation of posterior probabilities. Finally, efforts should be made in future studies to obtain more representative samples. These new samples should provide a better representation of the genetic diversity existing around the world, more precisely describing the populations of the native area and all the main invaded areas.

Acknowledgments We would like to thank Arnaud Gaigher and Pedro Barbosa for technical support, Douglas LeDoux and Julia Thompson for American sampling and Eric Lombaert for fruitful discussions. This work was funded by the EU REPHRAME Project (KBBE.2010.1.4-09).

References

Abelleira A, Picoaga A, Mansilla JP, Aguin O (2011) Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant Dis* 95:776

- Akbulut S, Linit MJ (1999) Flight performance of *Monochamus carolinensis* (Coleoptera: Cerambycidae) with respect to nematode phoresis and beetle characteristics. *Environ Entomol* 28:1014–1020
- Akbulut S, Stamps WT (2011) Insect vectors of the pinewood nematode: a review of the biology and ecology of *Monochamus* species. *For Pathol* 1–11
- Allendorf FW, Lundquist LL (2003) Introduction: population biology, evolution, and control of invasive species. *Conserv Biol* 17:24–30. doi:10.1046/j.1523-1739.2003.02365.x
- Barun A, Niemiller ML, Fitzpatrick BM, Fordyce JA, Simberloff D (2013) Can genetic data confirm or refute historical records? The island invasion of the small Indian mongoose (*Herpestes auropunctatus*). *Biol Invasions* 15:2243–2251. doi:10.1007/s10530-013-0447-6
- Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Montpellier (France), Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol* 19:2609–2625
- Bossdorf O, Auge H, Lafuma L, Rogers W, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144:1–11. doi:10.1007/s00442-005-0070-z
- Boucher AC, Mimee B, Montarry J, Bardou-Valette S, Belair G, Moffett P, Grenier E (2013) Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and determination of the origin of populations in Quebec. *Can Mol Phylogenet Evol* 69:75–82. doi:10.1016/j.ympev.2013.05.020
- Castagnone C, Abad P, Castagnone-Sereno P (2005) Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae). *Eur J Plant Pathol* 112:191–193. doi:10.1007/s10658-004-0580-2
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19:233–257
- Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24:621–631. doi:10.1093/molbev/msl191
- Cheng XY, Cheng FX, Xu RM, Xie BY (2008) Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: Aphelenchoididae) and its possible spread routes in China. *Heredity* 100:356–365. doi:10.1038/sj.hdy.6801082
- Cheng XY, Xie PZ, Cheng FX, Xu RM, Xie BY (2009) Competitive displacement of the native species *Bursaphelenchus mucronatus* by an alien species *Bursaphelenchus xylophilus* (Nematoda: Aphelenchida: Aphelenchoididae): a case of successful invasion. *Biol Invasions* 11:205–213
- Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A, Guillemaud T (2008) Invasion of Europe by the western corn rootworm,

- 806 *Diabrotica virgifera virgifera*: multiple transatlantic
807 introductions with various reductions of genetic diversity.
808 Mol Ecol 17:3614–3627
- 809 Dlugosch KM, Parker IM (2008) Invading populations of an
810 ornamental shrub show rapid life history evolution despite
811 genetic bottlenecks. Ecol Lett 11:701–709. doi:10.1111/j.
812 1461-0248.2008.01181.x
- 813 Dropkin VH, Foudin A, Kondo E, Linit MJ, Smith M, Robbins
814 K (1981) Pinewood nematode: a threat to US forests? Plant
815 Dis 65:1022–1027
- 816 Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a
817 website and program for visualizing STRUCTURE output
818 and implementing the Evanno method. Conserv Genet
819 Resour 4:359–361. doi:10.1007/s12686-011-9548-7
- 820 Estoup A, Guillemaud T (2010) Reconstructing routes of inva-
821 sion using genetic data: why, how and so what? Mol Ecol
822 19:4113–4130. doi:10.1111/j.1365-294X.2010.04773.x
- 823 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of
824 clusters of individuals using the software structure: a
825 simulation study. Mol Ecol 14:2611–2620
- 826 Evans HF, McNamara DG, Braasch H, Chadoeuf J, Magnusson
827 C (1996) Pest risk analysis (PRA) for the territories of
828 European Union (as PRA area) on *Bursaphelenchus xylo-*
829 *philus* and its vectors in the genus *Monochamus* OEPP/
830 EPPO. Bulletin 26:199–249
- 831 Facon B, Pointier JP, Glaubrecht M, Poux C, Jarne P, David P
832 (2003) A molecular phylogeography approach to biological
833 invasions of the New World by parthenogenetic Thiarid
834 snails. Mol Ecol 12:3027–3039. doi:10.1046/j.1365-294X.
835 2003.01972.x
- 836 Facon B, Genton BJ, Shykoff J, Jarne P, Estoup A, David P
837 (2006) A general eco-evolutionary framework for under-
838 standing bioinvasions. Trends Ecol Evol 21:130–135.
839 doi:10.1016/j.tree.2005.10.012
- 840 Falush D, Stephens M, Pritchard JK (2003) Inference of popu-
841 lation structure using multilocus genotype data: linked loci
842 and correlated allele frequencies. Genetics 164:1567–1587
- 843 Figueiredo J et al (2013) Assessment of the geographic origins
844 of pinewood nematode isolates via single nucleotide
845 polymorphism in effector genes. PLoS One 8:e83542.
846 doi:10.1371/journal.pone.0083542
- 847 Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG
848 (2012) What can DNA tell us about biological invasions?
849 Biol Invasions 14:245–253. doi:10.1007/s10530-011-
850 0064-1
- 851 Fonseca L et al (2012) The pinewood nematode, *Bursaphelen-*
852 *chus xylophilus*, in Madeira Island. Helminthologia
853 49:96–103
- 854 Fontaine MC, Gladieux P, Hood ME, Giraud T (2013) History
855 of the invasion of the anther smut pathogen on *Silene lat-*
856 *ifolia* in North America. New Phytol 198:946–956. doi:10.
857 1111/nph.12177
- 858 Frankham R (2005) Invasion biology—resolving the genetic
859 paradox in invasive species. Heredity 94:385. doi:10.1038/
860 sj.hdy.6800634
- 861 Futai K (2013) Pinewood nematode, *Bursaphelenchus xylo-*
862 *philus*. Annu Rev Phytopathol 51:61–83
- 863 Goudet J (2002) Fstat, a program to estimate and test gene
864 diversities and fixation indices. [http://www2.unil.ch/
865 popgen/softwares/fstat.htm](http://www2.unil.ch/popgen/softwares/fstat.htm)
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A
(2010) Inferring introduction routes of invasive species
using approximate Bayesian computation on microsatellite
data. Heredity 104:88–99. doi:10.1038/hdy.2009.92
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster
matching and permutation program for dealing with label
switching and multimodality in analysis of population
structure. Bioinformatics 23:1801–1806
- Kelager A, Pedersen J, Bruun H (2013) Multiple introductions
and no loss of genetic diversity: invasion history of Japa-
nese Rose, *Rosa rugosa*, in Europe. Biol Invasions
15:1125–1141. doi:10.1007/s10530-012-0356-0
- Keller SR, Taylor DR (2008) History, chance and adaptation
during biological invasion: separating stochastic pheno-
typic evolution from response to selection. Ecol Lett
11:852–866
- Kiritani K, Morimoto N (2004) Invasive insect and nematode
pests from North America. Glob Environ Res 8:75–88
- Kobayashi F, Yamane A, Ikeda T (1984) The Japanese pine
sawyer beetle as the vector of pine wilt disease. Annu Rev
Entomol 29:115–135
- Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos
JB (2004) Genetic variation increases during biological
invasion by a Cuban lizard. Nature 431:177–181. doi:10.
1038/nature02807
- Linit MJ (1988) Nematode–vector relationships in the pine wilt
disease system. J Nematol 20:227–235
- Mack RN, Simberloff D, Mark Lonsdale W, Evans H, Clout M,
Bazzaz FA (2000) Biotic invasions: causes, epidemiology,
global consequences, and control. Ecol Appl 10:689–710
- Mallez S et al (2013) First insights into the genetic diversity of
the pinewood nematode in its native area using new poly-
morphic microsatellite loci. PLoS One 8:e59165. doi:10.
1371/journal.pone.0059165
- Mamiya Y (1972) Pine wood nematode *Bursaphelenchus lig-*
nicolus Mamiya and Kiyohara, as a causal agent of pine
wilting disease. Rev Plant Protect Res 5:46–60
- Mamiya Y (1975) The life history of the pine wood nematode,
Bursaphelenchus lignicolus. Jpn J Nematol 5:16–25
- Mamiya Y (1976) Pine wilt disease caused by the pine wood
nematode, *Bursaphelenchus lignicolus*, in Japan. Jpn Agric
Res Q 10:206–211
- Mamiya Y (1983) Pathology of the pine wilt disease caused by
Bursaphelenchus xylophilus. Annu Rev Phytopathol
21:201–220. doi:10.1146/annurev.py.21.090183.001221
- Mamiya Y (1988) History of pine wilt disease in Japan. J Ne-
matol 20:219–226
- Mamiya Y (2004) Pine wilt disease in Japan. In: Mota M, Vieira
P (eds) The pinewood nematode, *Bursaphelenchus xylo-*
philus. Brill Academic, Leiden, pp 9–20
- Mamiya Y, Enda N (1972) Transmission of *Bursaphelenchus*
lignicolus (Nematoda: Aphelenchoididae) by *Monoch-*
amus alternatus (Coleoptera: Cerambycidae). Nematol-
ogica 18:159–162
- Metge K, Burgermeister W (2008) Analysis of *Bursaphelenchus*
xylophilus (Nematoda: Parasitaphelenchidae) provenances
using ISSR and RAPD fingerprints. In: Mota M, Vieira P
(eds) Pine wilt disease: a worldwide threat to forest eco-
systems. Springer, New York, pp 175–186. doi:10.1007/
978-1-4020-8455-3_15



- 926 Moon YS, Cheon HM, Lee S (2007) Occurrence of pine wilt
927 disease, caused by *Bursaphelenchus xylophilus*, from
928 *Pinus koraiensis* in Korea. J Nematol 39:87–88
- 929 Mota M, Braasch H, Bravo MA, Penas AC, Burgermeister W,
930 Metge K, Sousa E (1999) First report of *Bursaphelenchus*
931 *xylophilus* in Portugal and in Europe. Nematology
932 1:727–734
- 933 Naves PM, Camacho S, de Sousa EM, Quartau JA (2007)
934 Transmission of the pine wood nematode *Bursaphelenchus*
935 *xylophilus* through feeding activity of *Monochamus gal-*
936 *loprovincialis* (Col., Cerambycidae). J Appl Entomol
937 131:21–25. doi:10.1111/j.1439-0418.2006.01111.x
- 938 OEPP/EPPO (1997) Fiche informative sur les organismes de
939 quarantaine: *Bursaphelenchus xylophilus* no 158 OEPP/
940 EPPO Bulletin 1–12
- 941 Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic
942 assignment methods for the direct, real-time estimation of
943 migration rate: a simulation-based exploration of accuracy
944 and power. Mol Ecol 13:55–65. doi:10.1046/j.1365-294X.
945 2004.02008.x
- 946 Papura D, Burbán C, van Helden M, Giresse X, Nusillard B,
947 Guillemaud T, Kerdelhué C (2012) Microsatellite and
948 mitochondrial data provide evidence for a single major
949 introduction for the Nearctic leafhopper *Scaphoideus titanus*
950 in Europe. PLoS One 7:e36882. doi:10.1371/journal.pone.
951 0036882
- 952 Pascual M et al (2007) Introduction history of *Drosophila su-*
953 *bobscura* in the New World: a microsatellite-based survey
954 using ABC methods. Mol Ecol 16:3069–3083
- 955 Perdereau E, Bagnères AG, Bankhead-Dronnet S, Dupont S,
956 Zimmermann M, Vargo EL, Dedeine F (2013) Global
957 genetic analysis reveals the putative native source of the
958 invasive termite, *Reticulitermes flavipes*, in France. Mol
959 Ecol 22:1105–1119. doi:10.1111/mec.12140
- 960 Pereira F, Moreira C, Fonseca L, van Asch B, Mota M, Abrantes I,
961 Amorim A (2013) New insights into the phylogeny and
962 worldwide dispersion of two closely related nematode species,
963 *Bursaphelenchus xylophilus* and *Bursaphelenchus mucrona-*
964 *tus*. PLoS One 8. doi:10.1371/journal.pone.0056288
- 965 Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Es-
966 toup A (2004) GENECLASS2: a software for genetic
967 assignment and first-generation migrant detection. J Hered
968 95:536–539. doi:10.1093/jhered/esh074
- 969 Pritchard JK, Stephens M, Donnelly P (2000) Inference of
970 population structure using multilocus genotype data.
971 Genetics 155:945–959
- 972 Puth LM, Post DM (2005) Studying invasion: have we missed
973 the boat? Ecol Lett 8:715–721. doi:10.1111/j.1461-0248.
974 2005.00774.x
- 975 Rannala B, Mountain JL (1997) Detecting immigration by using
976 multilocus genotypes. Proc Natl Acad Sci 94:9197–9201
- 977 Raymond M, Rousset F (1995) An exact test for population
978 differentiation. Evolution 49:1280–1283. doi:10.2307/
979 2410454
- 980 Reed DH, Frankham R (2001) How closely correlated are
981 molecular and quantitative measures of genetic variation?
982 A meta-analysis. Evolution 55:1095–1103. doi:10.1111/j.
983 0014-3820.2001.tb00629.x
- 984 Reed DH, Frankham R (2003) Correlation between fitness and
985 genetic diversity. Conserv Biol 17:230–237. doi:10.1046/j.
986 1523-1739.2003.01236.x
- Robertson L, Cobacho Arcos S, Escuer M, Santiago Merino R,
Esparrago G, Abelleira A, Navas A (2011) Incidence of the
pinewood nematode *Bursaphelenchus xylophilus* Steiner &
Buhner, 1934 (Nickle, 1970) in Spain. Nematology
13:755–757
- Roman J, Darling JA (2007) Paradox lost: genetic diversity and
the success of aquatic invasions. Trends Ecol Evol
22:454–464. doi:10.1016/j.tree.2007.07.002
- Rosenberg NA (2004) DISTRUCT: a program for the graphical
display of population structure. Mol Ecol Notes 4:137–138.
doi:10.1046/j.1471-8286.2003.00566.x
- Rousset F (1997) Genetic differentiation and estimation of gene
flow from F-statistics under isolation by distance. Genetics
145:1219–1228
- Rousset F (2008) GENEPOP' 007: a complete re-implemen-
tation of the GENEPOP software for Windows and Linux.
Mol Ecol Resour 8:103–106. doi:10.1111/j.1471-8286.
2007.01931.x
- Saitou N, Nei M (1987) The neighbor-joining method: a new
method for reconstructing phylogenetic trees. Mol Biol
Evol 4:406–425
- Sakai AK et al (2001) The population biology of invasive spe-
cies. Annu Rev Ecol Syst 32:305–332. doi:10.1146/
annurev.ecolsys.32.081501.114037
- Shibata E (1986) Dispersal movement of the adult Japanese pine
sawyer, *Monochamus alternatus* hope (Coleoptera, Ce-
rambycidae) in a young pine forest. Appl Entomol Zool
21:184–186
- Simberloff D et al (2013) Impacts of biological invasions:
what's what and the way forward. Trends Ecol Evol (Per-
sonal edition) 28:58–66
- Sokal RR, Rohlf FJ (1995) Biometry The principles and practice
of statistics in biological research, 3rd edn. W.H. Freeman
and Company, New York, pp 236–242
- Soliman T, Mourits MCM, van der Werf W, Hengeveld GM,
Robinet C, Lansink A (2012) Framework for modelling
economic impacts of invasive species, applied to pine
wood nematode in Europe. PLoS One 7. doi:10.1371/
journal.pone.0045505
- Sousa E, Bravo MA, Pires J, Naves P, Penas AC, Bonifacio L,
Mota MM (2001) *Bursaphelenchus xylophilus* (Nematoda;
Aphelenchoididae) associated with *Monochamus gallo-*
provincialis (Coleoptera; Cerambycidae) in Portugal.
Nematology 3:89–91. doi:10.1163/156854101300106937
- Steiner G, Buhner EM (1934) *Aphelenchoides xylophilus*, N. SP.
A nematode associated with blue-stain and other fungi in
timber. J Agric Res 48:949–951
- Strong DR, Pemberton RW (2000) Biological control of
invading species-risk and reform. Science 288:1969–1970.
doi:10.1126/science.288.5473.1969
- Suzuki K (2002) Pine wilt disease—a threat to pine forest in
Europe. Dendrobiology 48:71–74
- Tares S, Abad P, Bruguier N, de Guiran G (1992) Identification
and evidence for relationships among geographical isolates
of *Bursaphelenchus* spp. (pinewood nematode) using
homologous DNA probes. Heredity 68:157–164
- Togashi K (1985) Transmission curves of *Bursaphelenchus*
xylophilus (Nematoda: Aphelenchoididae) from its vector,
Monochamus alternatus (Coleoptera: Cerambycidae), to
pine trees with reference to population performance. Appl
Entomol Zool 20:246–251

- Togashi K (1990) A field experiment on dispersal of newly emerged adults of *Monochamus alternatus* (Coleoptera: Cerambycidae). *Res Popul Ecol* 32:1–13
- Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proc Natl Acad Sci USA* 97:5948–5953. doi:10.1073/pnas.100110397
- Valadas V, Barbosa P, Espada M, Oliveira S, Mota M (2012a) The pinewood nematode, *Bursaphelenchus xylophilus*, in Portugal: possible introductions and spread routes of a serious biological invasion revealed by molecular methods. *Nematology* 00:1–13
- Valadas V, Laranjo M, Mota M, Oliveira S (2012a) Molecular characterization of Portuguese populations of the pinewood nematode *Bursaphelenchus xylophilus* using cytochrome *b* and cellulase genes. *J Helminthol*. doi:10.1017/S0022149X12000673
- van Wilgen BW, Moran VC, Hoffmann JH (2013) Some perspectives on the risks and benefits of biological control of invasive alien plants in the management of natural ecosystems. *Environ Manag* 52:531–540. doi:10.1007/s00267-013-0099-4
- Vicente C, Espada M, Vieira P, Mota M (2011) Pine wilt disease: a threat to European forestry. *Eur J Plant Pathol*. doi:10.1007/s10658-011-9924-x
- Vieira P, Burgermeister W, Mota M, Metge K, Silva G (2007) Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses. *J Nematol* 39:118–126
- Viglierchio DR, Schmitt RV (1983) On the methodology of nematode extraction from field samples: Baermann funnel modifications. *J Nematol* 15:438–444
- Vincent B, Altemayer V, Roux-Morabito G, Naves P, Sousa E, Lieutier F (2008) Competitive interaction between *Bursaphelenchus xylophilus* and the closely related species *Bursaphelenchus mucronatus*. *Nematology* 10:219–230. doi:10.1163/156854108783476403
- Wan XW, Liu YH, Zhang B (2012) Invasion history of the oriental fruit fly, *Bactrocera dorsalis*, in the Pacific-Asia Region: two main invasion routes. *PLoS One* 7. doi:10.1371/journal.pone.0036176
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annu Rev Ecol Evol Syst* 37:433–458. doi:10.1146/annurev.ecolsys.37.091305.110145
- Wilson JR, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends Ecol Evol* 24:136–144. doi:10.1016/j.tree.2008.10.007
- Wingfield MJ, Blanchette RA, Nicholls TH, Robbins K (1982) The pine wood nematode: a comparison of the situation in the United States and Japan. *Can J For Res* 12:71–75. doi:10.1139/x82-010
- Yoder M et al (2006) DESS : a versatile solution for preserving morphology and extractable DNA of nematodes. *Nematology* 8:367–376
- Zepeda-Paulo FA et al (2010) The invasion route for an insect pest species: the tobacco aphid in the New World. *Mol Ecol* 19:4738–4752. doi:10.1111/j.1365-294X.2010.04857.x
- Zhang K et al (2008) Molecular phylogeny of geographical isolates of *Bursaphelenchus xylophilus*: implications on the origin and spread of this species in China and worldwide. *J Nematol* 40:127–137
- Zhou ZH, Sakaue D, Wu BY, Hogetsu T (2007) Genetic structure of populations of the pinewood nematode *Bursaphelenchus xylophilus*, the pathogen of pine wilt disease, between and within pine forests. *Phytopathology* 97:304–310. doi:10.1094/phyto-97-3-0304

SUPPLEMENTARY INFORMATION

Table S1 Pairwise Weir & Cockerham estimates of F_{ST} (Weir & Cockerham 1984) between samples corrected for null alleles, implemented in FREENA (Chapuis & Estoup 2007).

Figure S1 Mean Log-likelihood and ΔK (Evanno et al. 2005) for each value of K tested in STRUCTURE analyses for (A) the USA, (B) Japan and (C) Portugal. The delta K from the Portuguese analysis is not shown, because a single cluster was inferred for in this area.

Figure S2 Bar plots of the coefficients of co-ancestry obtained with STRUCTURE for each value of K tested in the USA. Each bar corresponds to one individual nematode and each cluster is shown in a particular colour.

Figure S3 Weir & Cockerham estimates of F_{ST} corrected for null alleles (Chapuis & Estoup 2007; Weir & Cockerham 1984) between each Portuguese sample (except TR1, shown in Figure 3) and each Japanese (in green) or American (in orange) sample. The asterisk and the dashed lines indicate the sample presenting the lowest F_{ST} value and the mean F_{ST} value across samples for each area, respectively.

Figure S4 Mean Log-likelihood of the multilocus individual assignment ($L_{i \rightarrow s}$) of each Portuguese sample (except TR1, shown in Fig.3) to each Japanese (in green) or American (in orange) sample. The asterisk and the dashed lines indicate the sample presenting the maximum $L_{i \rightarrow s}$ and the mean $L_{i \rightarrow s}$ across samples for each area, respectively.

Figure S5 NJ tree based on the Cavalli-Sforza & Edwards distances (Cavalli-Sforza & Edwards 1967) corrected for null alleles, obtained with FREENA (Chapuis & Estoup 2007).

Table S1

	MO1	MO2	NE1	NE2	NE5	NE6	NE9	NE10	NE12	NE13b	NE14	NE15	NE19	NE22	NE23	NE24	V19	MA1	Mad23PC	Mad24C	E182	128S	E1069	AM2	Comporta	TR1	TR2	Jap308	Jap212	Jap120	Kasumig2	Kasumig3	Kasumig5
MO2	0.06																																
NE1	0.29	0.27																															
NE2	0.25	0.19	0.19																														
NE5	0.28	0.24	0.12	0.15																													
NE6	0.27	0.26	0.17	0.16	0.14																												
NE9	0.23	0.19	0.14	0.11	0.09	0.14																											
NE10	0.33	0.29	0.18	0.25	0.09	0.28	0.15																										
NE12	0.39	0.40	0.49	0.50	0.47	0.50	0.41	0.52																									
NE13b	0.17	0.17	0.22	0.23	0.19	0.24	0.12	0.19	0.29																								
NE14	0.22	0.20	0.23	0.19	0.15	0.15	0.11	0.21	0.42	0.12																							
NE15	0.31	0.26	0.23	0.24	0.20	0.37	0.17	0.15	0.60	0.23	0.27																						
NE19	0.19	0.15	0.32	0.31	0.29	0.38	0.29	0.34	0.44	0.24	0.32	0.35																					
NE22	0.25	0.21	0.22	0.11	0.18	0.20	0.12	0.28	0.40	0.19	0.14	0.29	0.31																				
NE23	0.41	0.39	0.50	0.45	0.41	0.50	0.43	0.42	0.66	0.47	0.42	0.57	0.48	0.48																			
NE24	0.50	0.50	0.62	0.64	0.57	0.63	0.50	0.55	0.70	0.55	0.54	0.61	0.57	0.61	0.59																		
V19	0.33	0.30	0.50	0.47	0.48	0.50	0.39	0.46	0.69	0.45	0.34	0.45	0.49	0.43	0.57	0.71																	
MA1	0.43	0.39	0.60	0.55	0.58	0.57	0.48	0.61	0.61	0.50	0.49	0.60	0.54	0.51	0.72	0.76	0.60																
Mad23PC	0.49	0.45	0.61	0.41	0.57	0.56	0.38	0.61	0.77	0.58	0.51	0.59	0.62	0.44	0.82	0.90	0.74	0.80															
Mad24C	0.45	0.40	0.55	0.35	0.51	0.52	0.34	0.57	0.74	0.52	0.47	0.54	0.57	0.38	0.80	0.88	0.70	0.77	0.00														
E182	0.49	0.46	0.61	0.42	0.58	0.57	0.39	0.62	0.77	0.59	0.52	0.60	0.63	0.45	0.82	0.90	0.74	0.81	0.00	0.00													
128S	0.48	0.44	0.58	0.35	0.54	0.52	0.36	0.59	0.76	0.57	0.49	0.57	0.62	0.39	0.79	0.88	0.71	0.80	0.29	0.24	0.30												
E1069	0.50	0.47	0.62	0.43	0.59	0.57	0.39	0.62	0.77	0.59	0.52	0.60	0.64	0.45	0.82	0.90	0.74	0.81	0.00	0.00	0.00	0.31											
AM2	0.54	0.52	0.68	0.50	0.65	0.62	0.44	0.66	0.81	0.65	0.57	0.65	0.69	0.51	0.84	0.92	0.78	0.84	0.00	0.00	0.00	0.36	0.00										
Comporta	0.57	0.56	0.71	0.54	0.69	0.65	0.47	0.69	0.83	0.69	0.60	0.68	0.72	0.55	0.86	0.93	0.80	0.86	0.00	0.00	0.00	0.40	0.00	0.00									
TR1	0.58	0.57	0.72	0.56	0.71	0.66	0.48	0.70	0.83	0.70	0.61	0.69	0.73	0.57	0.87	0.94	0.81	0.87	0.00	0.00	0.00	0.41	0.00	0.00	0.00								
TR2	0.57	0.56	0.71	0.54	0.69	0.65	0.47	0.69	0.82	0.68	0.60	0.68	0.72	0.55	0.86	0.93	0.80	0.86	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00							
Jap308	0.56	0.57	0.67	0.61	0.72	0.70	0.51	0.70	0.82	0.68	0.61	0.68	0.69	0.57	0.85	0.93	0.79	0.85	0.99	0.99	0.99	0.91	0.99	0.99	0.99	0.99	0.99						
Jap212	0.57	0.58	0.68	0.63	0.73	0.70	0.52	0.71	0.82	0.69	0.62	0.69	0.70	0.58	0.86	0.93	0.79	0.86	0.99	0.99	0.99	0.92	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.00			
Jap120	0.55	0.56	0.66	0.60	0.71	0.69	0.50	0.70	0.81	0.67	0.60	0.68	0.68	0.55	0.85	0.93	0.78	0.85	0.99	0.99	0.99	0.91	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.00	0.00		
Kasumig2	0.59	0.62	0.74	0.55	0.73	0.68	0.50	0.72	0.84	0.72	0.63	0.71	0.77	0.61	0.86	0.94	0.81	0.86	0.99	0.99	0.99	0.91	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99		
Kasumig3	0.44	0.43	0.61	0.43	0.56	0.51	0.43	0.62	0.67	0.56	0.54	0.63	0.59	0.45	0.71	0.83	0.75	0.76	0.74	0.72	0.75	0.69	0.75	0.78	0.80	0.81	0.80	0.88	0.89	0.88	0.86		
Kasumig5	0.53	0.59	0.79	0.61	0.76	0.72	0.57	0.74	0.82	0.71	0.65	0.75	0.72	0.64	0.82	0.93	0.83	0.86	0.99	0.99	0.99	0.94	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.83	
Kosa	0.53	0.53	0.61	0.42	0.57	0.56	0.39	0.62	0.77	0.61	0.52	0.58	0.66	0.48	0.75	0.86	0.70	0.75	0.73	0.70	0.73	0.67	0.73	0.76	0.78	0.79	0.78	0.82	0.83	0.82	0.63	0.75	0.79

Figure S1

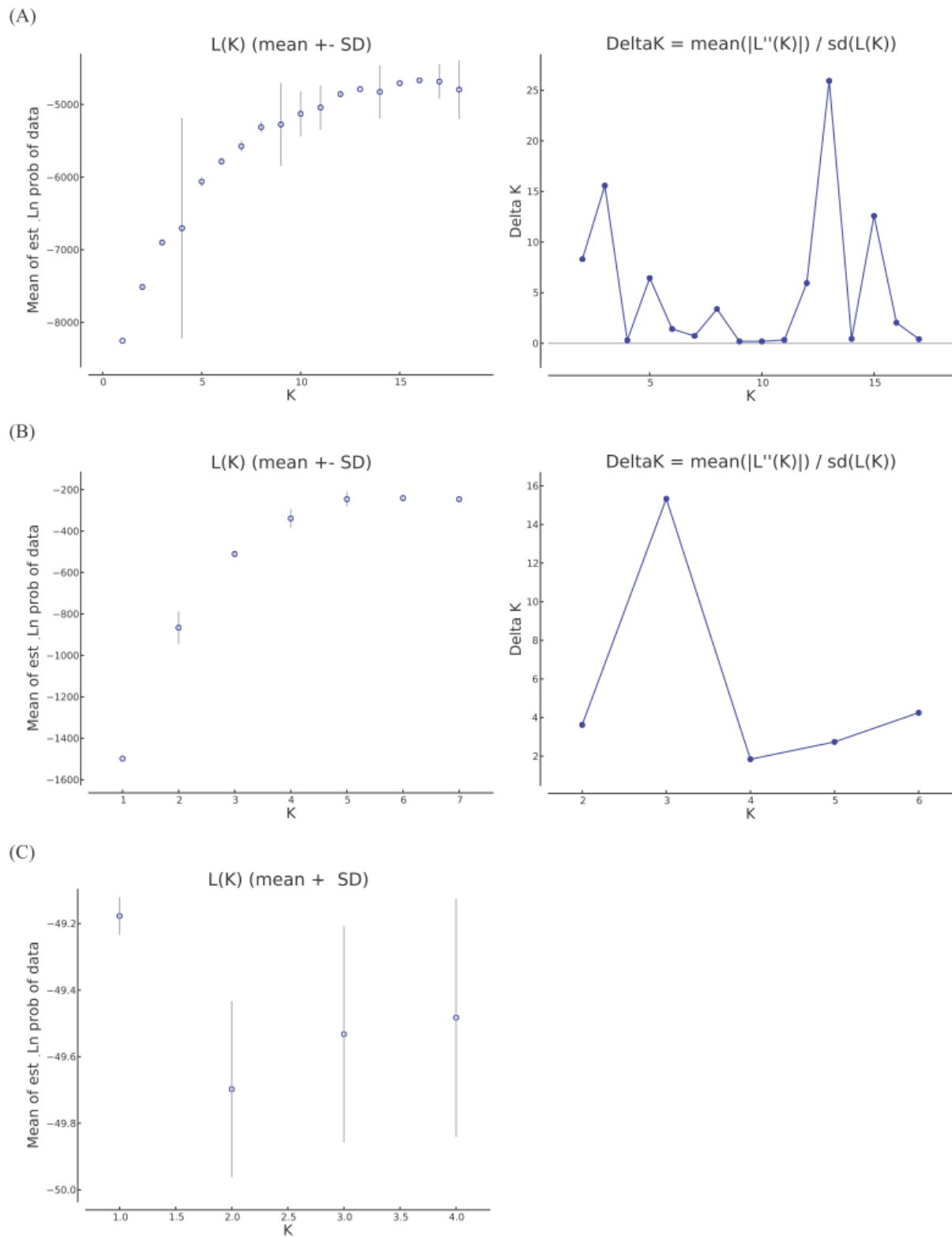


Figure S2

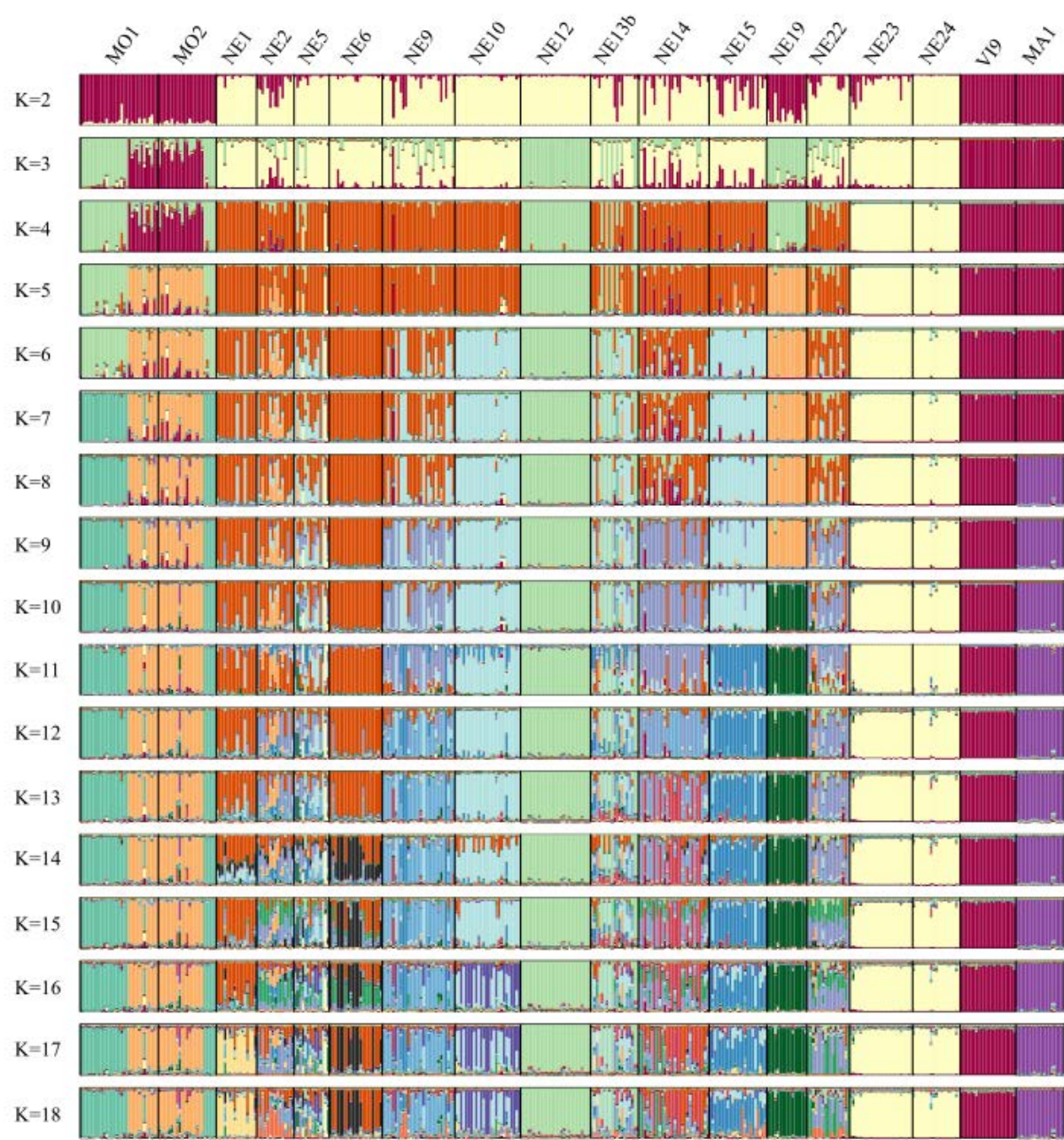


Figure S3

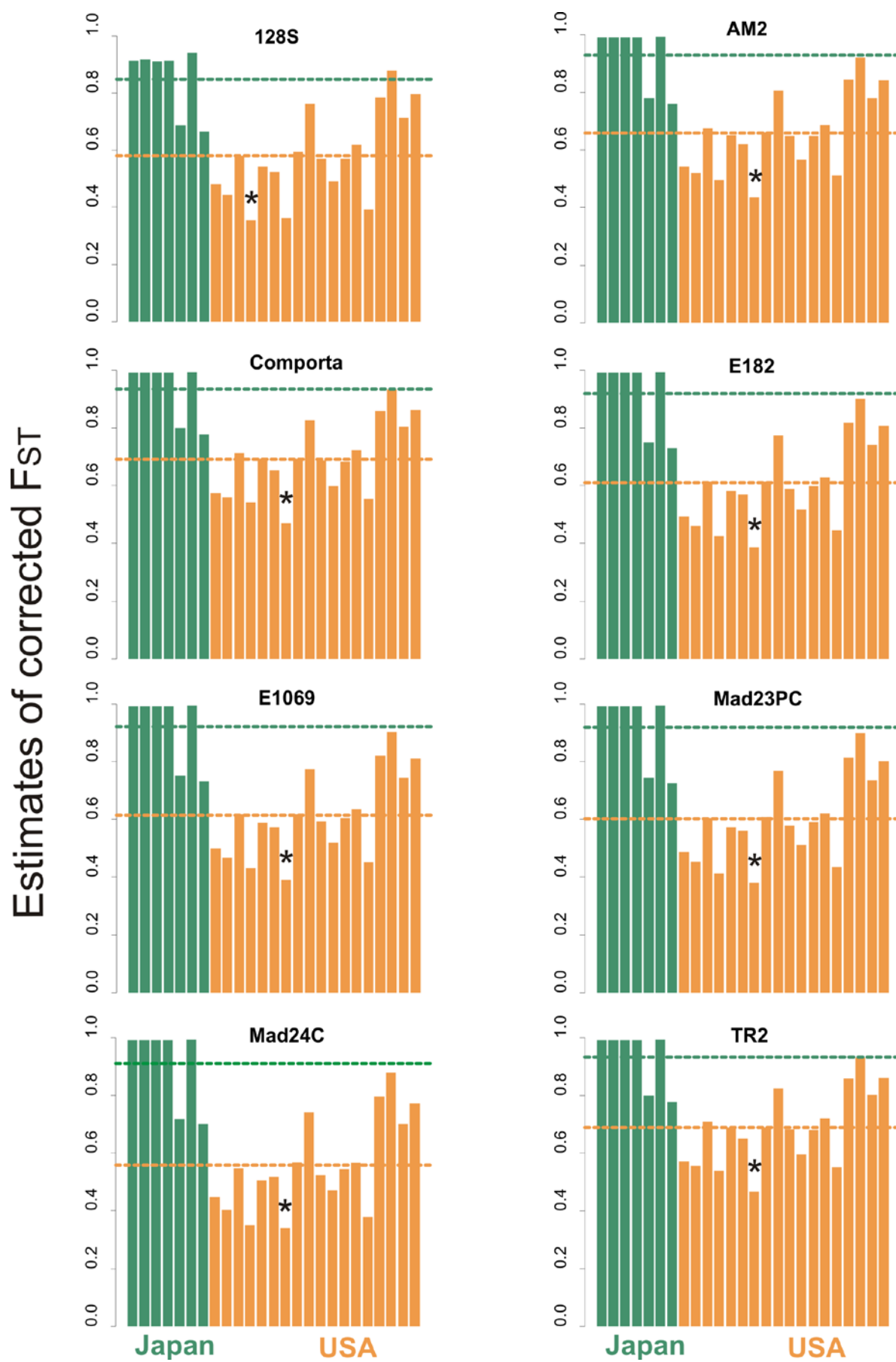


Figure S4

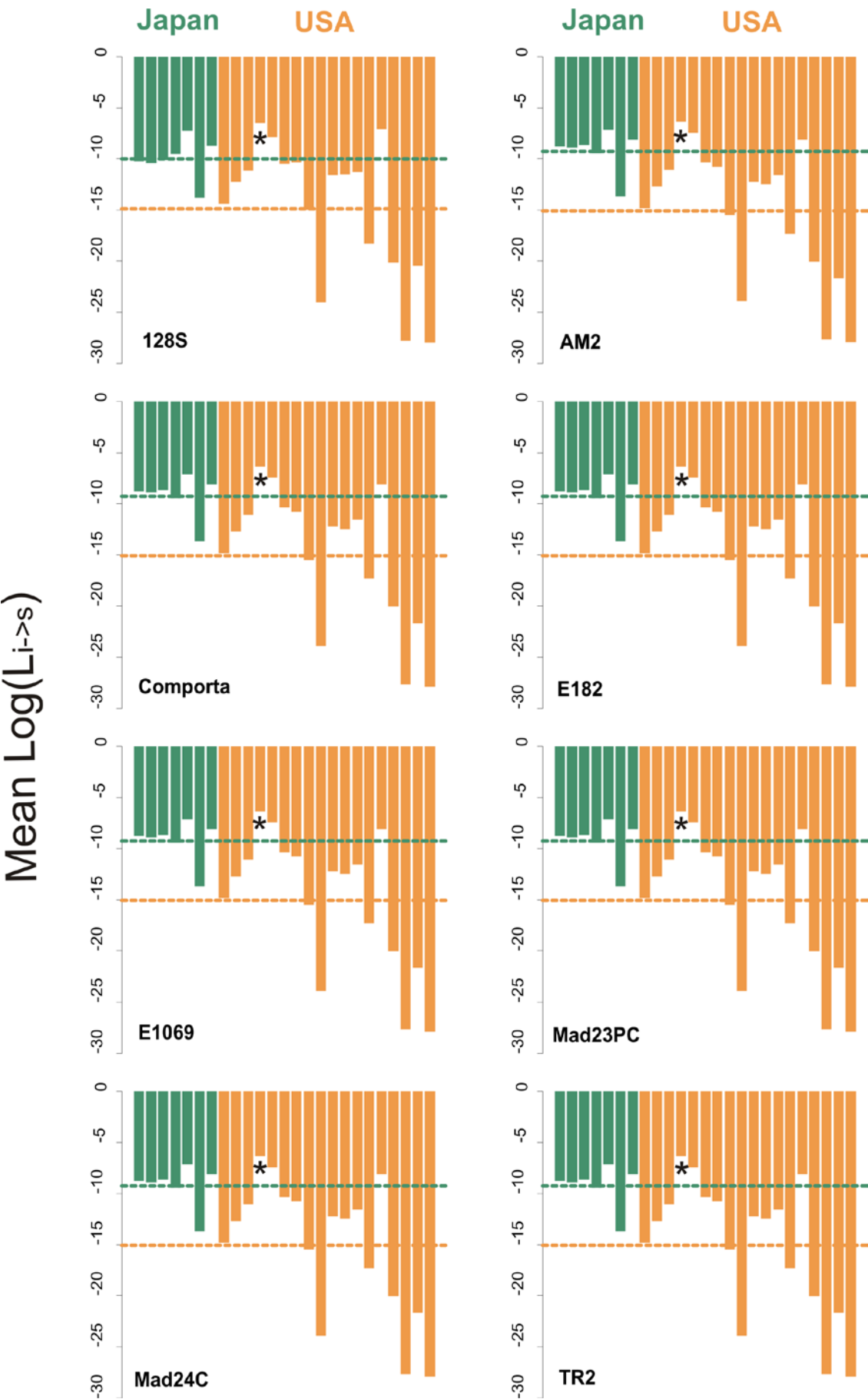
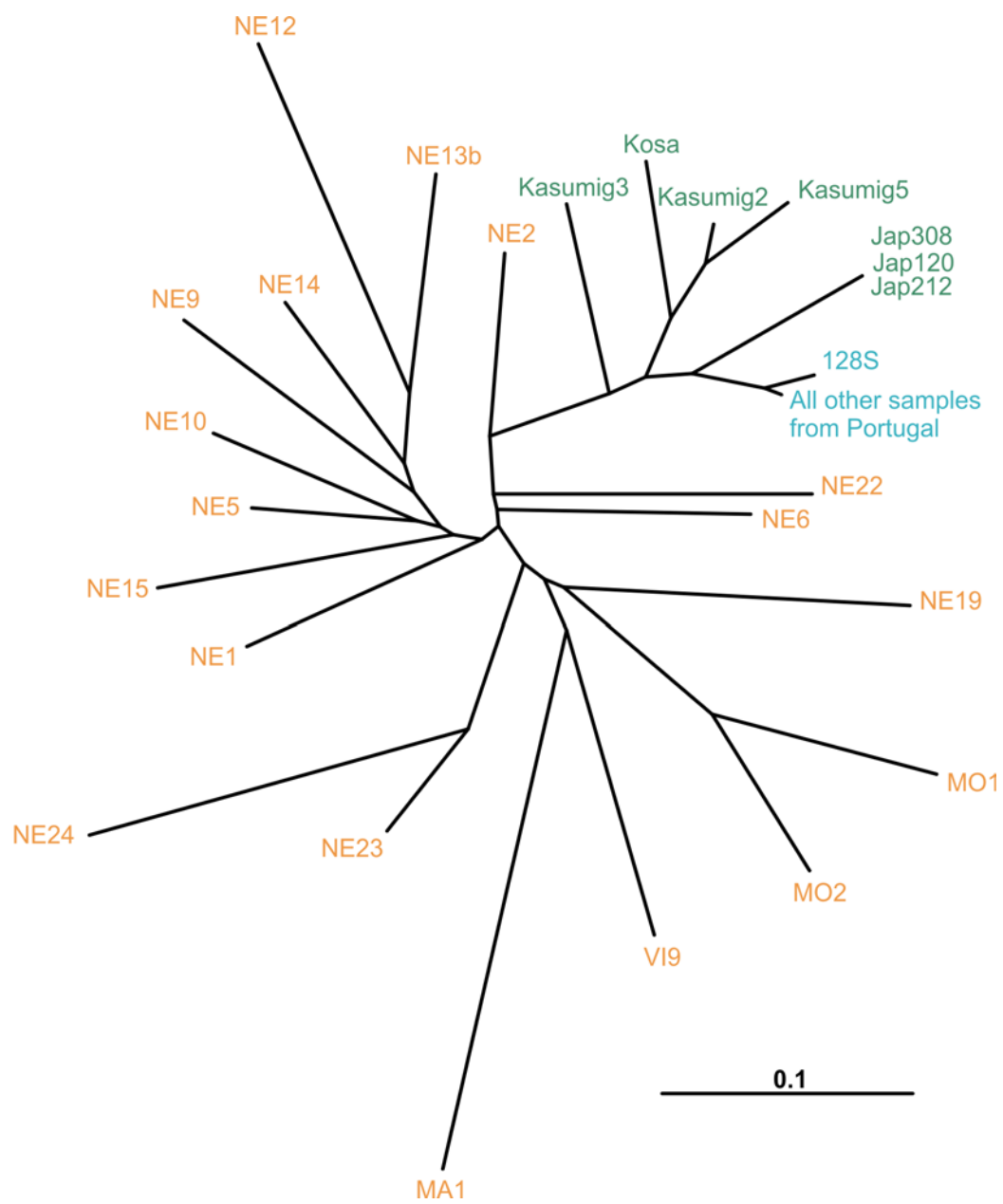


Figure S5



2. Effet "collection"

L'observation du manque de diversité génétique au sein des souches de collection, ainsi que les résultats obtenus lors de la mise au point des marqueurs microsatellites nous ont poussé à approfondir les différences qu'il pouvait exister entre les souches de collection et les échantillons de terrain. Le maintien au laboratoire d'échantillons de populations naturelles modifie potentiellement l'intégrité génétique des échantillons. Il nous a donc semblé important d'étudier cela dans le cas précis du PWN, dans la mesure où beaucoup d'études se font sur des souches de collection.

Cette étude se rapporte à l'article suivant :

- Mallez S, Castagnone C, Morel-Journel T, Guillemaud T, Castagnone-Sereno P. (2014). "Collection Effect" in the pinewood nematode: is it worth maintaining strains under laboratory conditions ? Soumis à Evolutionary Applications.

Elle est résumée ci-dessous et la version soumise de l'article est présentée à la suite du résumé.

❖ Contexte et objectifs

La possibilité de maintenir des espèces/individus au laboratoire ou en captivité est un atout majeur pour le contrôle biologique, la recherche scientifique ou la conservation. Cependant, ce maintien implique des changements importants pour les espèces maintenues. Ces changements sont d'ordre environnemental et démographique et peuvent provoquer des différences phénotypiques et génétiques entre les populations naturelles et les souches de collection/populations captives. Plusieurs études ont déjà été conduites pour analyser les conséquences du maintien au laboratoire et la plupart d'entre elles ont mis en évidence des

différences génétiques entre les populations naturelles et les populations de laboratoire (Arias et al. 2005; Coe et al. 2009; Fiumera et al. 2000; Nowak et al. 2007; Stohler et al. 2004).

Le PWN s'élève très bien sur champignon au laboratoire et l'échantillonnage de populations naturelles est compliqué. Beaucoup d'études réalisées sur le PWN se font donc sur des souches de collection, éventuellement maintenues au laboratoire depuis un grand nombre de générations. Cependant, peu de choses sont connues sur le niveau de diversité génétique des souches de collection par rapport aux échantillons de populations naturelles et sur le niveau de différenciation génétique qu'il peut exister entre ces catégories.

Dans cette étude, nous avons donc comparé la diversité génétique et la différenciation génétique entre des souches de collection et des échantillons de populations naturelles de l'aire native, en utilisant les marqueurs microsatellites développés précédemment.

❖ Matériel et méthodes

Nous avons analysé 28 échantillons de populations naturelles et 12 souches de laboratoire de l'aire native. Toutes les souches de laboratoire ne sont pas maintenues depuis le même nombre d'années (et donc de générations) au laboratoire. Nous avons donc divisé les souches de collection en deux catégories : les souches récentes, âgées de 5 ans ou moins, et les souches anciennes, âgées de plus de 20 ans (voir Tableau 1 dans l'article). Trois catégories ont été obtenues et comparées : "Terrain", "Récentes" et "Anciennes".

Dans un premier temps, la diversité génétique des différentes catégories a été comparée *via* le nombre d'allèles et l'hétérozygotie attendue, et la différenciation génétique entre paires de catégories *via* le F_{st} (Weir & Cockerham 1984) et le D de Jost (Jost 2008). Deux types de tests ont été utilisés pour réaliser les comparaisons en fonction du nombre d'observations par catégorie : des tests de permutations, lorsqu'au moins une des catégories comparée présentait un petit nombre d'observations, et des tests de Wilcoxon dans les autres

cas. Tous les tests ont été réalisés par marqueur et la méthode généralisée de Fisher due à Lancaster (Chen 2011) a été utilisée pour combiner les probabilités obtenues par marqueur.

Dans un second temps, nous avons simulé *in silico* le maintien d'une souche au laboratoire pour évaluer l'effet de la durée de ce maintien sur la diversité génétique de la souche. Le maintien d'une souche au laboratoire nécessite des événements de repiquages. Il y a donc alternance de périodes de repiquage et de périodes de croissance. C'est l'alternance de ces différentes périodes que nous avons simulée (voir Figure 1 dans l'article). Plus précisément, nous avons simulé, grâce au logiciel DYABC (Cornuet et al. 2014), une souche de collection initiée à partir d'un échantillon de population naturelle et maintenue pendant 15 ans au laboratoire. Nous avons ensuite "échantillonné" cette souche chaque année pour suivre l'évolution temporelle de sa diversité génétique. N'ayant pas d'informations précises sur la taille efficace des souches de collection, nous avons testé différentes gammes de tailles efficaces. Cela nous permettait également d'évaluer une large gamme d'intensité de dérive génétique. Nous avons enfin exploré l'impact d'accidents d'élevage qui peuvent avoir lieu lors du maintien pendant de nombreuses années sur la diversité génétique d'une souche de laboratoire.

❖ Résultats

Nous avons mis en évidence des différences génétiques significatives entre les différentes catégories d'échantillons. Le sens de ces différences est apparu dépendant de l'âge des souches de collection. Ainsi les "vieilles" souches de laboratoire se sont avérées significativement moins diversifiées que les souches récentes et les échantillons de terrain. Les souches récentes sont quant à elle apparues légèrement plus diversifiées que les échantillons de terrain. Concernant le niveau de différenciation, les vieilles souches sont

apparues génétiquement plus différenciées de n'importe quelle autre catégorie que les souches récentes et les échantillons de terrain.

Les simulations montrent une augmentation de la diversité génétique, que ce soit du nombre d'allèles ou de l'hétérozygotie attendue, au cours du temps pour quasiment toutes les tailles efficaces testées. Nous avons enfin montré que les accidents avaient un effet important sur le niveau de diversité génétique en le réduisant de manière conséquente.

❖ Discussion

Nous avons observé une relation non monotone entre la diversité génétique des échantillons et le temps passé au laboratoire : les échantillons de populations naturelles étant moins diversifiés que les souches récentes de laboratoire mais plus diversifiés que les souches anciennes. L'augmentation de la diversité génétique observée dans les souches récentes est supportée par les résultats des simulations et peut s'expliquer par une simplification du cycle de vie du PWN au laboratoire. En conditions naturelles, le PWN fait face à plusieurs défis pour réussir à se reproduire et à se propager, passant par de petites tailles efficaces qui se traduisent par des diversités intra-population très faibles. Ces défis n'existant plus au laboratoire, les tailles efficaces du PWN sont probablement plus grandes et la diversité augmente avec le temps au grès des mutations. De plus, les boîtes de Pétri, dans lesquelles les nématodes sont maintenus, correspondent à des environnements spatialement plus petits que les arbres et permettent sûrement un meilleur brassage des nématodes, résultant en une plus faible variance inter-individuelle de reproduction et une plus faible dérive. En ce qui concerne la faible diversité des souches anciennes, l'hypothèse la plus probable est la survenue d'accidents qui peuvent soudainement et de manière importante réduire la taille efficace d'une souche et donc sa diversité génétique (comme montré par les simulations). Cette hypothèse est d'autant plus probable que le maintien au laboratoire est long. Bien que nous ayons avancé

des hypothèses pour expliquer l'augmentation de la diversité génétique au cours du temps pendant le maintien en collection, ce résultat n'était pas du tout attendu et nous a surpris. Nous nous attendions en effet à ce que toutes les souches de collection soit moins diversifiées que les échantillons de populations naturelles, comme cela est couramment montré lorsque des comparaisons populations naturelles – souches de collection sont effectuées (Arias et al. 2005; Bech et al. 2010; Ciosi et al. 2014; Coe et al. 2009; Fiumera et al. 2000; Norris et al. 2001; Nowak et al. 2007; Stohler et al. 2004).

Nous avons observé une relation plus simple et attendue entre le niveau de différenciation entre échantillons et la durée du maintien au laboratoire, à savoir une corrélation positive. Cela peut s'expliquer par l'action combinée de la dérive génétique, de l'accumulation de mutations dans chaque souche et de l'absence de l'absence de flux de gènes entre les souches due à leur confinement.

Au-delà des différences génétiques explorées ici, le maintien au laboratoire peut également favoriser des traits d'histoire de vie différents comparés aux conditions naturelles, étant donné que les pressions de sélection sont très probablement différentes entre laboratoire et nature. Cela est tout à fait envisageable chez le PWN dans la mesure où son cycle de vie naturellement complexe n'est pas reproduit au laboratoire.

Malgré la nécessité de réaliser des études supplémentaires pour clarifier l'effet "collection" chez cette espèce, les analyses effectuées dans cette étude illustrent déjà les différences génétiques qui peuvent exister entre les échantillons de populations naturelles et les souches de laboratoire, soulignant l'importance d'être prudent lorsque l'on interprète des résultats obtenus grâce à des souches de laboratoire.

❖ Au-delà de l'article...

Nous avons utilisé la méthode généralisée de Fisher proposée par Lancaster (Chen 2011) pour combiner les probabilités obtenues par marqueur et évaluer l'existence d'une tendance générale. À notre connaissance, cette méthode est peu courante et nous trouvons donc utile de justifier notre choix. La méthode la plus usuelle en génétique des populations pour combiner les probabilités est la méthode de Fisher. Cette méthode est ainsi très souvent utilisée pour évaluer si les informations obtenues à partir de différents tests, ayant la même hypothèse nulle, permettent ou non de rejeter cette hypothèse nulle partagée (Rice 1990; Whitlock 2005). Cependant, cette méthode s'avère très sensible aux probabilités faibles, voire très faibles, et ne serait donc pas appropriée pour cela (Rice 1990). Elle serait plus adéquate pour tester si au moins l'un des tests combinés est significatif. Le résultat après application de la méthode de combinaison de Fisher peut ainsi induire en erreur lorsque l'on souhaite évaluer une tendance générale. C'est pourquoi nous avons opté pour une méthode alternative, la méthode généralisée de Fisher proposée par Lancaster, qui n'est pas sensible aux probabilités faibles (Chen 2011). À noter cependant, que la méthode généralisée de Fisher repose sur les quantiles de la distribution du χ^2 . Les probabilités de 1 sont donc problématiques pour cette méthode, puisque leurs quantiles renvoient à l'infini.

Article n°3

Mallez S, Castagnone C, Morel-Journel T, Guillemaud T, Castagnone-Sereno P. (2014). "Collection Effect" in the pinewood nematode: is it worth maintaining strains under laboratory conditions ? Soumis à Evolutionary Applications.

Cet article a été soumis à Evolutionary Applications le 16 juillet 2014.

"Collection Effect" in the pinewood nematode: is it worth maintaining strains under laboratory conditions?"

Sophie Mallez^{1,2,3}, Chantal Castagnone^{1,2,3}, Thibaut Morel-Journal^{1,2,3}, Thomas Guillemaud^{1,2,3}, Philippe Castagnone-Sereno^{1,2,3}

¹ INRA, UMR 1355 Institut Sophia Agrobiotech, 06903 Sophia Antipolis, France

² Université de Nice Sophia Antipolis, UMR Institut Sophia Agrobiotech, 06903 Sophia Antipolis cedex, France

³ CNRS, UMR 7254 Institut Sophia Agrobiotech, 06903 Sophia Antipolis cedex, France

Corresponding Author:

Sophie Mallez

Email: sophie.mallez@sophia.inra.fr

Tel: (33) 492386424

ABSTRACT

The maintenance of species in laboratory (or in captivity) implies environmental and demographic changes compared to natural conditions, which may result in phenotypic and genetic differences between natural populations and laboratory strains (or captive populations). Here, we compared the genetic variability and genetic differentiation, using 16 microsatellite markers, between laboratory strains and field samples of the pinewood nematode (PWN), *Bursaphelenchus xylophilus*, an invasive species responsible of the pine wilt disease. Our statistical analyses demonstrated significant genetic differences between laboratory strains and field samples: (i) recent laboratory strains genetically more and old laboratory strains genetically less diversified than field samples and (ii) old laboratory strains more differentiated with whatever other samples than field samples and recent laboratory strains. The non monotonous relationship between genetic variation and duration of the maintenance was supported by computer simulations. We then discussed on the interpretation of results obtained from laboratory strains and the relevance of their extrapolation to natural populations.

Keywords: *Bursaphelenchus xylophilus*; microsatellite; genetic diversity; polymorphism; coalescence; simulations; differentiation.

INTRODUCTION

Many species are maintained in laboratory or in captivity for several purposes, such as biological control, scientific research or conservation. This maintenance may imply changes for the maintained species because rearing or laboratory conditions are generally different from natural ones. First, environmental differences may exist between natural and laboratory/captivity conditions, which must be adapted to suit with laboratory/captivity constraints and correspond to the final objectives of maintenance. Laboratory/captivity conditions are usually stable and controlled, without natural enemies, environmental stochasticity and heterogeneity. Consequently, selection pressures may dramatically be modified, some of them being relaxed and others magnified. This may have strong impact on species features *via* a decrease of adaptation capacity, because of a reduced genetic variability (Masel, King, and Maughan 2007) and *via* the adaptation to laboratory/captivity, resulting in changes in behavior (Carducci and Jakob 2000; Kelley, Magurran, and Macías-García 2005; De Gregorio et al. 2013; but see Kuriwada et al. 2014). Second, the demography and the dynamics of populations may also be disturbed in laboratory/captivity. For instance, laboratory strains (defined as a sample of a natural population maintained for several generations in the laboratory) are generally isolated from other laboratory strains or natural populations: they are not integrated in a metapopulation dynamics and thus not prone to gene flow. As a consequence, there is no natural input of genetic diversity once a strain or population is established in the laboratory or in captivity, except by mutation. Only artificial input of genetic diversity may exist and can be planned (for instance in the case of scientific experiments or conservation programs) or fortuitous (in case of contaminations). Moreover, the establishment and maintenance of laboratory strains and captive populations may provoke founder effects and bottlenecks, reducing their genetic diversity and their effective size (Nei, Maruyama, and Chakraborty 1975), i.e. the size that would have the population if it were at mutation-drift equilibrium (Wright 1931). A limited number of individuals may generate genetic drift that in turn reduces genetic variability and increase the probability of inbreeding (Frankham 2005). This may have detrimental consequences on fitness (Reed and Frankham 2003) and on adaptive potential (Willi, Van Buskirk, and Hoffmann 2006), *via* mechanisms such as inbreeding depression (Frankham and Ralls 1998; Woodworth et al. 2002) or fixation of deleterious mutations.

Consequently, laboratory strains and captive populations evolve in particular environmental and demographic conditions and several studies have thus investigated the

consequences of these particularities by comparing laboratory strains and field samples (Stuart and Gaugler 1996; Lanzaro et al. 1998; Barata et al. 2000; Fiumera, Parker, and Fuerst 2000; Norris et al. 2001; Stohler, Curtis, and Minchella 2004; Arias et al. 2005; Nowak et al. 2007; Coe et al. 2009; Bech et al. 2010; Ciosi, Masiga, and Turner 2014). A majority of these studies demonstrated a loss of genetic diversity in laboratory strains compared to field samples in several species such as, for example, human pathogens (Stohler, Curtis, and Minchella 2004; Bech et al. 2010), vectors of diseases (Norris et al. 2001; Arias et al. 2005; Ciosi, Masiga, and Turner 2014), endangered species (Fiumera, Parker, and Fuerst 2000) and model organisms (insects in Nowak et al. 2007; fishes in Coe et al. 2009). Also, these studies highlighted the implications of these differences on results of experimental tests such as ecotoxicological tests (Barata et al. 2000; Nowak et al. 2007; Coe et al. 2009), control techniques such as the sterile insect technique (Ciosi, Masiga, and Turner 2014) and trials of vaccine efficacy (Stohler, Curtis, and Minchella 2004).

Focusing on laboratory maintenance, we were interested in the pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle 1970; Nematoda: Aphelenchoididae). It is an invasive species responsible of the pine wilt disease (Mamiya 1972, 1976; Futai 2013 for a review) and thus it is the subject of many studies worldwide. Most studies on this species were performed on laboratory strains and many research laboratories around the world possess their own for various reasons. First, it is a good candidate for the maintenance of laboratory collection because (i) it is a small organism (less than 1µm), which requires limited space for storage of strains, (ii) it is easily grown in Petri dishes on fungi (e.g. *Botrytis cinerea*) and (iii) it displays a short life cycle (4-5 days at 25°C), resulting in a rapid population growth. Second, working with natural populations requires the sampling of PWN in the wild which remains a difficult task because it is almost impossible to detect infected trees by eye. Pieces of wood must first be sampled and then be subject to a laborious protocol to determine nematode infection and extract individuals. Third, the PWN is a species of strict quarantine concern, which precludes any experiment to be conducted *in natura* in countries from which the pest has not been reported. Many studies were thus carried out on laboratory strains of PWN: pathogenicity/virulence tests (Mota et al. 2006; Zhu et al. 2012), researches on control strategies (Oh et al. 2009), behavior/biology (Wang et al. 2005; Son et al. 2010) and genetic studies (Tares et al. 1992; Mota et al. 2006; Valadas et al. 2012; Figueiredo et al. 2013). Laboratory strains have often been established for a long time (up to more than 20 years, corresponding to more than 500 generations in standard laboratory conditions) but to our knowledge, little attention has been given to the genetic integrity or

variability of these strains. Therefore, little is known about the level of genetic diversity of laboratory strains compared to field samples of the PWN and about the genetic differentiation that may exist between those categories of samples.

The purpose of this study was thus to compare the genetic variability and to measure the level of differentiation between laboratory strains and field samples of PWN populations. For this, we genotyped laboratory strains and natural samples originating from the native area, i.e., North America, of this nematode, using 16 microsatellite markers previously characterized in this species (Mallez et al. 2013).

MATERIAL AND METHODS

Natural life cycle and rearing conditions

In *natura*, the biological cycle of the PWN is complex and involves at least two partners: the pine tree and the insect vector (Evans et al. 1996). The PWN and its vector are closely associated and their life cycles closely match (Mamiya 1972), resulting for the PWN in alternating periods of dispersion/transmission and reproduction into the tree. In laboratory, the PWN is reared in Petri dishes on fungi (e.g., *Botrytis cinerea*) and its life cycle is thus extremely simplified. The life cycle depends mainly on the rearing temperature, which determines its duration (Mamiya 1975). It also relies on recurrent transplantation of nematodes from old dishes to new ones, necessary to ensure resource availability and survival of PWN strains for long-term maintenance.

Nematode samples and microsatellite loci genotyping

Twenty eight field samples and twelve laboratory strains of PWN from the native area were analysed in this study. Their characteristics are listed in Table 1. All the field samples were extracted from wood samples that were collected directly from field locations and were grouped under the category “Field” for the analyses. Laboratory strains were aged from 5 to 30 years and were divided into two categories: “Recent” for laboratory strains aged 5 years or less and “Old” for laboratory strains aged more than 20 years (see Table 1). No permission was required to collect samples of PWN from the infested areas and we obtained an official agreement from the French authorities (#2012060-0004) for the importation and manipulation

of this quarantine organism at Institut Sophia Agrobiotech facilities. The DNA was extracted by thermal shock and 16 microsatellite markers were amplified on single nematodes as described in Mallez et al. (2013). Genotype scoring was performed using an ABI 3700 sequencer (Applied Biosystems) with the 500 LIZTM GeneScanTM size standard (Applied Biosystems) and GenemarkerTM version 1.75 software (SoftGenetics LLC).

Data analysis

Description of the sample diversity and differentiation

For each sample, we determined the maximum (among loci) of the number of alleles per locus (Max Na) detected over all loci, the mean number of alleles (Mean Na) per sample and the mean expected heterozygosity (Mean He) per sample with FSTAT version 2.9.3.2 (Goudet 2002). We also tested the hypothesis of uniform genotype frequencies between samples by Fisher's exact test (Raymond and Rousset 1995) implemented in GENEPOP (Rousset 2008). As non-independent multiple tests were performed, sequential Bonferroni correction (Sokal and Rohlf 1995) was carried out to adjust significance levels. We calculated Weir & Cockerham estimates of F_{ST} (Weir and Cockerham 1984) between samples, corrected for null alleles with FREENA (Chapuis and Estoup 2007) due to the presence of null alleles in some of the markers (see Mallez & al., submitted). As F_{ST} is dependent on the within-population variability (Charlesworth 1998; Hedrick 1999; Jost 2008; Gerlach et al. 2010), we also calculated the D_{est} (Jost 2008) to assess differentiation between samples, with the SMOGD software (Crawford 2010).

Comparisons between the "Field", "Recent" and "Old" categories

To compare the genetic diversity (Na and He) between the three categories defined above (i.e. "Field", "Recent" and "Old") and the level of differentiation (corrected F_{ST} and D_{est}) between pairs of categories, we carried out two different types of tests depending on the available number of observations per category. Permutation tests were used when the number of observations was small in at least one category compared; Wilcoxon tests were used in other comparisons. In permutation tests, random combinations of observed values are generated and then are compared to the observed combination. In this study, all possible combinations were used when feasible. It was not the case only for the comparison between differentiation within "Field" vs. differentiation within "Recent" for which 200,000,000 permutations were used. The difference between means observed in each category was computed and compared to the

distribution of differences between means obtained from permutations to assess significance of the permutation tests. All the comparisons were performed at the locus level and the generalized Fisher method due to Lancaster (Chen 2011) was used for combining probabilities from locus-level tests. All statistical analyses were performed with R version 3.0.1 (R Development Core Team 2012).

Finally, we built neighbour joining (NJ) trees (Saitou and Nei 1987) with POPULATION software version 1.2.30 (<http://bioinformatics.org/~tryphon/populations/>), based on Cavalli-Sforza & Edwards genetic distances (Cavalli-Sforza and Edwards 1967) with bootstraps on locus obtained from 2,000 permutations and on genetic distances Cavalli-Sforza & Edwards (Cavalli-Sforza and Edwards 1967) corrected for null alleles with FREENA (Chapuis and Estoup 2007), but for which bootstrap supports cannot be obtained with the available tools.

Simulation of the laboratory maintenance

The maintenance of strains requires transplanting nematodes from old Petri dishes to new ones, resulting in the alternation of periods of inoculation and periods of growth/proliferation. More precisely, in the laboratory, for each transplanting event, we inoculate new Petri dishes with about 1,000-1,500 nematodes. After three weeks at 25°C, the new dishes exhibit well-developed populations (more than hundreds of thousands nematodes) and are kept at 15°C for approximately five months to slow down the growth of the PWN (1 generation in 12 days at this temperature). Thus, two transplanting events are carried out each year for each strain in collection. Using the coalescent theory (Nordborg 2007) and the DIYABC software version 2.0 (Cornuet et al. 2014), we simulated the effect of laboratory maintenance — genetic drift due to limited population size of the PWN in Petri dishes — of a strain on its genetic diversity (see Figure 1). In a few words, we simulated a laboratory strain that was sampled in the field 15 years ago. We then simulated samples taken from this laboratory strain each year to follow the temporal dynamics of its genetic diversity. Here, we simulated genetic drift due to limited effective population size thanks to the coalescence theory (Nordborg 2007). It allows the simulation of a sample of gene copies subjected to coalescence (the merge of two or more gene copies into a common ancestor backward in time), and mutation events (forward in time once the coalescence topology is obtained). For each of the 16 microsatellite loci, the genetic diversity of 50 diploid individuals (100 copies of genes per locus) sampled each year from the collection strain (i.e., after two transplanting events) was simulated. Backward simulations of each set of gene copy lineages encompassing a maximum of 15 years of genetic drift were

performed to build 1,000 genetic data sets per condition. The mean number of alleles and the mean expected heterozygosity (over the 16 microsatellite markers) was computed. Three effective population sizes were used for the simulations: N_i , the effective population size in the field; N_p , the effective population size at inoculation and N_g , the effective population size during PWN growth in Petri dishes (see Figure 1). Given the exponential growth of PWN, the period of inoculation was thought to be quite short and was estimated to last 2 generations (between 8-10 days). The period of proliferation/growth was fixed to 16 generations completing one event of transplanting (just over 5 months, taking into account the change in temperature). The distribution of N_i was chosen to match the genetic diversity encountered in the field (see Table 1) and corresponded to a log-uniform distribution between 10-1,000. As described earlier, we transplant about 1,000-1,500 individuals and the population size reaches hundreds of thousands individuals. Consequently, for the two other effective population sizes (N_p and N_g), uniform distributions with different ranges of values were tested in order to use a large scale of genetic drift intensity: [10-50] and [100-500] for N_p ; and [100-500], [1,000-5,000] and [10,000-50,000] for N_g . It is worth noting that the effective population size is normally lower than the census population size because it is the size that would have an ideal population at equilibrium in which genetic drift and mutation would result in the observed genetic variability (Wright 1931). We used a generalized stepwise mutation model to simulate mutations at the microsatellite markers in a forward process (Estoup, Jarne, and Cornuet 2002) and with the default mutation parameters in DIYABC version 2.0 (Cornuet et al. 2014). We finally explored the impact of rearing accidents that may occur during long-term maintenance on the genetic diversity of a strain. For that, we simulated a drastic reduction in N_p , i.e. uniform distribution between 2 and 10 during 6 generations, to represent a problem during one transplanting event.

RESULTS

Description of the sample diversity and differentiation

The level of genetic diversity was heterogeneous among “Field” samples. The mean N_a per sample ranged from 1.25 (in NE23) to 3.94 (in KS1) and the mean H_e ranged from 0.07 (in NE23 and in NE24) to 0.46 (in KS1, see Table 1). In laboratory strains, the level of genetic diversity was more homogeneous within the “Recent” and the “Old” categories: the mean N_a

per strains ranged from 1 (in US9) to 3.69 (in USDE5) and the mean H_e ranged from 0 (in US9) to 0.35 (in USDE5), with the “Recent” strains regrouping the higher values (see Table 1). A maximum number of 5 or more alleles was found in 15 out of the 28 “Field” samples (with a maximum of 10 alleles in KS1) and in all “Recent” strains (with a maximum of 7 alleles in USDE5). All the “Old” strains displayed less than 4 alleles (with 4 alleles found in US2, BC and St William; see Table 1).

Overall, the samples were significantly genetically differentiated (Fisher's exact test, $p < 10^{-3}$). Only one test between two “Field” samples was only marginally significant after correction (NC2 vs GA4, Fisher's exact test, $p = 0.0493$). The estimates of F_{ST} corrected (or not, data not shown) for null alleles reached very high values and ranged from 0.018 (NC2 vs USDE5) to 0.933 (US9 vs Q52A; for more details see Table S1). The D_{est} reached also very high values and ranged from 0.002 (NC2 vs USDE5) to 0.875 (US10 vs NE24; for more details see Table S1).

Comparison between "Field", "Recent" and "Old" categories

We observed significant differences in mean N_a and mean H_e between the three categories (see Figure 2). As expected, the “Old” category was less diversified than the “Field” category according to the mean N_a (one-sided permutation tests with 30,260,340 combinations, Lancaster procedure, mean N_a “Old” = 1.51 vs mean N_a “Field” = 2.23, $p < 10^{-10}$) and the mean H_e (one-sided permutation tests with 30,260,340 combinations, Lancaster procedure, mean H_e “Old” = 0.13 vs mean H_e “Field” = 0.27, $p < 10^{-10}$). The “Recent” category also exhibited a mean N_a (one-sided permutation tests with 495 combinations, Lancaster procedure, mean N_a “Recent” = 2.69 vs mean N_a “Old” = 1.51, $p < 10^{-10}$) and a mean H_e (one-sided permutation tests with 495 combinations, Lancaster procedure, mean H_e “Recent” = 0.31 vs mean H_e “Old” = 0.13, $p < 10^{-10}$) significantly higher than the “Old” category. Surprisingly, the mean N_a and the mean H_e of the “Recent” category appeared to be slightly larger than those of the “Field” category for most markers (see Figure 2). These differences were globally significant for both mean N_a (one-sided permutation tests with 35960 combinations, Lancaster procedure, mean N_a “Recent” = 2.69 vs mean N_a “Field” = 2.23, $p = 4.84 \times 10^{-7}$) and mean H_e (one-sided permutation tests with 35960 combinations, Lancaster procedure, mean H_e “Recent” = 0.31 vs mean H_e “Field” = 0.27, $p = 0.009$).

We observed significant differences of levels of differentiation between pairs of categories and within each category with the corrected estimate of F_{ST} and the D_{est} (see Figures 3 and 4).

The differentiation between "Field" and "Old" was significantly larger than the differentiation between "Field" and "Recent" categories for the two statistics (one-sided Wilcoxon tests, Lancaster procedure, $F_{ST} = 0.45$ and 0.16 and $D_{est} = 0.52$ and 0.23 for "Field-Old" and "Field-Recent" comparisons respectively, $p < 10^{-10}$). Similarly, the differentiation between "Recent" and "Old" categories was significantly larger than between "Field" and "Recent" categories for the two statistics (one-sided Wilcoxon tests, Lancaster procedure, $F_{ST} = 0.40$ and 0.16 and $D_{est} = 0.49$ and 0.23 for "Recent-Old" and "Field-Recent" comparisons respectively, $p < 10^{-10}$). The differentiation between "Field" and "Old" categories was also significantly larger than between "Recent" and "Old" categories for the corrected estimate of F_{ST} (one-sided Wilcoxon tests, Lancaster procedure, "Field-Old" = 0.45 vs "Recent-Old" = 0.40 , $p = 10^{-10}$) but there was no significant difference observed with the D_{est} (two-sided Wilcoxon tests, Lancaster procedure, "Field-Old" = 0.52 vs "Recent-Old" = 0.49 , $p = 0.57$). Within each category, the "Old" one exhibited the largest level of differentiation compared to the "Field" (one-sided Wilcoxon tests, Lancaster procedure, $F_{ST} = 0.52$ and 0.19 and $D_{est} = 0.58$ and 0.25 for "Old" and "Field" comparisons respectively, $p < 10^{-10}$) and the "Recent" (one-sided permutation tests with 1,344,904 combinations, Lancaster procedure, $F_{ST} = 0.52$ and 0.07 and $D_{est} = 0.58$ and 0.11 for "Old" and "Recent" comparisons respectively, $p < 10^{-10}$) categories for the corrected estimate of F_{ST} and for the D_{est} . The level of differentiation within the "Field" category was significantly larger than within the "Recent" category (one-sided permutation tests with 200,000,000 permutations, Lancaster procedure, $F_{ST} = 0.19$ and 0.07 and $D_{est} = 0.25$ and 0.11 for "Field" and "Recent" comparisons respectively, $p < 10^{-10}$). Finally, in the NJ trees, we observed that the majority of the "Old" strains were more distant from the other samples than the "Recent" strains (one-sided Wilcoxon test applied on the pairwise Cavalli-Sforza & Edward distances, $p < 10^{-10}$), which were all grouped with the "Field" samples (see Figure S1).

Simulation of the laboratory maintenance

We simulated the collection transplanting protocol using different combinations of distributions for N_p and N_g to evaluate their effects on the genetic diversity of the strain in collection (see Figures 6A and 6B). We observed that only one combination resulted in a stable genetic diversity along collection maintenance, of the same order of magnitude as that of the field. This combination corresponded to a mean N_p of 30 (unif [10-50]) and a mean N_g of 300 (unif [100-500]). All the other combinations resulted in an increase in genetic diversity

compared to the field. The increase was even more important when the distributions of the parameters included the largest values. Also, we observed that the differences were more pronounced between different N_p (for the same N_g) than between different N_g (for the same N_p). In other words, a reduction in N_p (for constant N_g) resulted in a larger reduction of the genetic diversity than a reduction in N_g (for constant N_p). We observed that rearing accidents have an important effect by dramatically reducing the genetic diversity, either the mean number of alleles or the mean expected heterozygosity. The genetic diversity did not recover completely after the accident and remained lower than that without accident (see Figure 5).

DISCUSSION

In this study, we compared field samples and laboratory strains based on their genetic diversity and their level of differentiation to evaluate if and to what extent laboratory strains are genetically different from natural populations. It appeared that the laboratory strains were significantly different from the field samples for these two aspects. However, the relationship between the extent of genetic differences and the duration of the maintenance in collection was not straightforward.

The genetic diversity comparisons gave variable results depending on the age of the laboratory strains. The “Old” strains (more than 20 years in collection) appeared significantly less genetically diversified than “Field” samples for both N_a and H_e , as already observed in previous studies on other biological models (Stohler, Curtis, and Minchella 2004; Nowak et al. 2007; Coe et al. 2009). In contrast and surprisingly, the “Recent” strains (5 years or less in collection) appeared significantly more genetically diversified than “Field” samples for both N_a and H_e . Considering the large variance of the genetic diversity found in the field, one might assume that all the “Old” strains were originally less genetically diversified and be part of the lower range encountered in the field, and inversely for the “Recent” strains. However, assuming that all samples were collected randomly, this would very unlikely lead to statistically significant results. In addition, the coalescent simulations corroborated the increased genetic diversity in “Recent” strains compared to “Field” samples, thus raising two questions:

(i) How to explain that the setting in collection tends to increase the genetic diversity in PWN strains?

(ii) Why, in this case, did the “Old” strains appear to be less genetically diversified than the “Field” samples and *a fortiori* than the “Recent” strains?

Concerning the first question, several hypotheses may be proposed to explain why genetic diversity tends to increase in the first years of collection. Here, it is important to remember that genetic diversity depends on the effective population size (Wright 1931). At mutation drift equilibrium, the genetic diversity is positively linked to theta, the product of the mutation rate and the effective population size. The life cycle of the PWN is complex *in natura* (Futai 2013) where the nematode must overcome several demographic challenges to reproduce and propagate, that go through low population size. These challenges do not exist anymore in laboratory where the PWN grows only on fungus in a controlled environment that allows very large census population sizes. The artificial simplification of the life cycle may thus limit the occurrence of demographic bottlenecks and lower the impact of the genetic drift, which is predominant in the field, as demonstrated by Mallez et al. (submitted). In line with this simplification, it is thus probable that the effective population sizes are larger in the laboratory than in the field and thus may account for an increase of genetic diversity. It is noteworthy that the only population size parameters (N_p and N_g) keeping the genetic diversity constant, i.e. at a level equivalent to that of the field, are very small. They are unrealistic compared to the census population sizes that are routinely reached by in laboratory strains (more than hundreds of thousands individuals). The Petri dishes, where the PWN is reared, finally constitute very favorable environments. They are spatially smaller compared to the tree, allowing the nematodes to mix and reproduce all together more easily, avoiding problems of mate limitation.

The extremely low genetic diversity detected in "Old" strains and the absence of significant difference between different samples from the same strain (i.e., 2 sampling points separated by 15 years, data not shown) contradict the results of the simulations and raise the second question mentioned above. The most likely hypothesis to explain this discrepancy is the occurrence of collection accidents (e.g., invasion of the strains by mites, falls of Petri dishes, problems in the control of incubator temperature...), which may suddenly and dramatically reduce the effective size of strains. These accidents are probably occurring and may even be recurrent when the collection maintenance is long. Simulations agree with this hypothesis and illustrate the impact of accidents on the genetic diversity of a strain. Moreover, they show that inoculation periods represent very crucial steps, which have the most pronounced influence on the genetic diversity: lowering N_p dramatically reduced the genetic diversity.

Besides the level of genetic diversity, the levels of genetic differentiation appeared also different between laboratory strains and field samples. The corrected estimate of F_{ST} and the D_{est} gave in most cases concordant results. The genetic differentiation between "Field" and "Old" and between "Recent" and "Old" were significantly larger than the differentiation between "Field" and "Recent". Then, the level of differentiation within the "Old" category was also significantly higher than within the "Field" and the "Recent" categories. The level of genetic differentiation between samples thus appears positively correlated with the duration of laboratory maintenance: the older the laboratory rearing of a strain, the larger the differentiation with whatever other samples. This can be explained by the genetic drift coupled with the absence of gene flow due to the confinement of the laboratory strains, as suggested by Simões et al. (2010). Different strains can move toward fixation of different alleles because genetic drift is a random phenomenon. Moreover, this phenomenon is more pronounced in confined and isolated populations, where the genetic drift is not compensated by migration, thereby increasing the differentiation between different laboratory strains and between laboratory strains and natural populations.

However, further investigations are needed to clarify the results obtained in this study. Although we simulated the temporal dynamics of a PWN strain genetic diversity, we analysed real strains from a single sampling point only. Consequently, we have no information about the stability or temporal fluctuation of genetic diversity along the collection maintenance. An experimental monitoring of laboratory strains from their establishment to their maintenance in the long term, as carried out for another species of worms in Bech et al. (2010), would be of precious help to disentangle the sense of the collection effect in *B. xylophilus* and determine how the genetic diversity evolves during collection maintenance.

Working with neutral markers, we could only reveal global genetic features of laboratory strains during laboratory maintenance compared to natural populations. Indeed, the use of microsatellite markers gives insights into the global variability of entire genomes that mostly depends on demographic factors (e.g., effective population sizes, gene flow, although reproduction mode or mutation rate may also have an impact). Importantly, it does not give access to the genetic variability of particular genomic regions that may be the target of specific selection pressures. Although differences in global genomic variability may reflect global differences in the capacity to adapt or to respond to environmental stress or pressures (Reed and Frankham 2003; but see Hoffman, Forcada, and Amos 2006; Willi, Van Buskirk, and Hoffmann 2006; Dlugosch and Parker 2008), our study is facing the problem of the representativeness of the results, as already addressed in other studies (Stohler, Curtis, and

Minchella 2004; Nowak et al. 2007; Coe et al. 2009; Ciosi, Masiga, and Turner 2014). Stuart and Gaugler (1996) highlighted the problem of the adaptation to laboratory conditions in the entomopathogenic nematode species *Steinernema glaseri* and its consequences on life-history traits, such as the level of infectivity, the reproductive potential or the sex-ratio. Indeed, the selection pressures may completely change in laboratory compared to natural conditions and it is reasonable to expect that selection will act differently and favour different traits in laboratory. This may be particularly true for the PWN, where strains are reared on fungi and its life cycle in laboratory is completely different from what occurs in natural conditions. The complex life cycle of the PWN (Futai 2013 for a review) is not identically reproduced in laboratory and this can generate differences, other than the level of genetic diversity. Further studies trying to document this aspect of laboratory maintenance on the PWN life-history traits are needed to precisely evaluate the value of PWN laboratory strains for experimental purposes.

As a conclusion, our results illustrate the differences that may exist between laboratory strains and natural populations of PWN. It is thus very important to genetically characterise the samples before their setting in collection because it provides useful information to tune the transplanting protocol. Maintaining a constant genetic diversity (at the level of the field) in collection is possible but implies a special attention to the parameters of the transplanting, i.e. the effective population sizes. Laboratory collections may constitute very useful tools (Frankham 1999; Houle and Rowe 2003), mostly when obtaining/manipulating samples from the field is difficult. However, it is crucial to keep in mind that their use may not be appropriate for all scientific questions. Results must always be interpreted with caution when obtained from laboratory strains and the relevance of their extrapolation to natural populations must be carefully evaluated.

DATA ARCHIVING STATEMENT

Data for this study are available at: to be completed after manuscript is accepted for publication.

REFERENCES

- Arias, L., E. E. Bejarano, E. Marquez, J. Moncada, I. Velez, and S. Uribe. 2005. Mitochondrial DNA divergence between wild and laboratory populations of *Anopheles albimanus* Wiedemann (Diptera : Culicidae). *Neotropical Entomology* 34 (3):499-506.
- Barata, C., D. J. Baird, F. Amat, and A. Soares. 2000. Comparing population response to contaminants between laboratory and field: an approach using *Daphnia magna* ephippial egg banks. *Functional Ecology* 14 (4):513-523.
- Bech, N., S. Beltran, J. Portela, A. Rognon, J. F. Allienne, J. Boissier, and A. Théron. 2010. Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory. *Infection, Genetics and Evolution* 10 (7):1039-1045.
- Carducci, J. P. and E. M. Jakob. 2000. Rearing environment affects behaviour of jumping spiders. *Animal Behaviour* 59 (1):39-46.
- Cavalli-Sforza, L. L. and A. W. F. Edwards. 1967. Phylogenetic analysis. Models and estimation procedures. *The American Journal of Humans Genetics* 19:233 - 257.
- Chapuis, M. P. and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24 (3):621-631.
- Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* 15 (5):538-543.
- Chen, Z. 2011. Is the weighted z-test the best method for combining probabilities from independent tests? *Journal of Evolutionary Biology* 24 (4):926-930.
- Ciosi, M., D. K. Masiga, and C. M. R. Turner. 2014. Laboratory Colonisation and Genetic Bottlenecks in the Tsetse Fly *Glossina pallidipes*. *PLoS Negl Trop Dis* 8 (2):e2697.
- Coe, T. S., P. B. Hamilton, A. M. Griffiths, D. J. Hodgson, M. A. Wahab, and C. R. Tyler. 2009. Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies. *Ecotoxicology* 18 (1):144-150.
- Cornuet, J. M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J. M. Marin, and A. Estoup. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30 (8):1187-1189.
- Crawford, N. G. 2010. SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources* 10 (3):556-557.
- De Gregorio, B. A., P. J. Weatherhead, T. D. Tuberville, and J. H. Sperry. 2013. Time in captivity affects foraging behavior of ratsnakes: implications for translocation. *Herpetological Conservation and Biology* 8 (3):581-590.
- Dlugosch, K. M. and I. M. Parker. 2008. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters* 11 (7):701-709.
- Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11 (9):1591-1604.
- Evans, H. F., D. G. McNamara, H. Braasch, J. Chadoeuf, and C. Magnusson. 1996. Pest Risk Analysis (PRA) for the territories of European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *OEPP/EPPO Bulletin* 26 (2):199 - 249.
- Figueiredo, J., M. J. Simões, P. Gomes, C. Barroso, D. Pinho, L. Conceição, L. Fonseca, I. Abrantes, M. Pinheiro, and C. Egas. 2013. Assessment of the Geographic Origins of

- Pinewood Nematode Isolates via Single Nucleotide Polymorphism in Effector Genes. *PLoS One* 8 (12):e83542.
- Fiumera, A. C., P. G. Parker, and P. A. Fuerst. 2000. Effective Population Size and Maintenance of Genetic Diversity in Captive-Bred Populations of a Lake Victoria Cichlid. *Conservation Biology* 14 (3):886-892.
- Frankham, R. 1999. Resolving conceptual issues in conservation genetics: the roles of laboratory species and meta-analyses *Hereditas* 130:195 - 201.
- Frankham, R. and K. Ralls. 1998. Conservation biology - Inbreeding leads to extinction. *Nature* 392 (6675):441-442.
- Frankham, R. 2005. Genetics and extinction. *Biological Conservation* 126 (2):131-140.
- Futai, K. 2013. Pinewood nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 51:61-83.
- Gerlach, G., A. Jueterbock, P. Kraemer, J. Deppermann, and P. Harmand. 2010. Calculations of population differentiation based on G_{st} and D: forget G_{st} but not all of statistics! *Molecular Ecology* 19 (18):3845-3852.
- Goudet, J. 2002. Fstat, a program to estimate and test gene diversities and fixation indices. Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Hedrick, P. W. 1999. Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. *Evolution* 53 (2):313-318.
- Hoffman, J. I., J. Forcada, and W. Amos. 2006. No relationship between microsatellite variation and neonatal fitness in Antarctic fur seals, *Arctocephalus gazella*. *Molecular Ecology* 15 (7):1995-2005.
- Houle, D. and L. Rowe. 2003. Natural Selection in a Bottle. *The American Naturalist* 161 (1):50-67.
- Jost, L. 2008. G(ST) and its relatives do not measure differentiation. *Molecular Ecology* 17 (18):4015-4026.
- Kelley, J. L., A. E. Magurran, and C. Macías-García. 2005. The influence of rearing experience on the behaviour of an endangered Mexican fish, *Skiffia multipunctata*. *Biological Conservation* 122 (2):223-230.
- Kuriwada, T., N. Kumano, K. Shiromoto, and D. Haraguchi. 2014. Mass-rearing conditions do not affect responsiveness to sex pheromone and flight activity in sweetpotato weevils. *Journal of Applied Entomology* 138 (4):254-259.
- Lanzaro, G. C., B. Alexander, J. P. Mutebi, J. Montoya-Lerma, and A. Warburg. 1998. Genetic variation among natural and laboratory colony populations of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera : Psychodidae) from Colombia. *Memorias Do Instituto Oswaldo Cruz* 93 (1):65-69.
- Mallez, S., C. Castagnone, M. Espada, P. Vieira, J. D. Eisenback, M. Mota, T. Guillemaud, and P. Castagnone-Sereno. 2013. First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. *PLoS One* 8 (3):e59165.
- Mallez, S., C. Castagnone, M. Espada, P. Vieira, J. D. Eisenback, M. Harrell, M. Mota, T. Aikawa, M. Akiba, H. Kosaka, P. Castagnone-Sereno, and T. Guillemaud. Worldwide invasion routes of the pinewood nematode: what can we infer from population genetics analyses? *Submitted to Biological Invasions*.
- Mamiya, Y. 1972. Pine wood nematode *Bursaphelenchus lignicolus* Mamiya and Kiyohara, as a causal agent of pine wilting disease. *Review of Plant Protection Research* 5:46 - 60.
- . 1975. The life history of the pine wood nematode, *Bursaphelenchus lignicolus*. *Japanese Journal of Nematology* 5:16 - 25.

- . 1976. Pine wilt disease caused by the pine wood nematode, *Bursaphelenchus lignicolus*, in Japan. *Japan Agricultural Research Quarterly* 10 (4):206 - 211.
- Masel, J., O. D. King, and H. Maughan. 2007. The loss of adaptive plasticity during long periods of environmental stasis. *The American Naturalist* 169:38 - 46.
- Mota, M. M., S. Takemoto, Y. Takeuchi, N. Hara, and K. Futai. 2006. Comparative studies between Portuguese and Japanese isolates of the pinewood nematode, *Bursaphelenchus xylophilus*. *Journal of Nematology* 38 (4):429 - 433.
- Nei, M., T. Maruyama, and R. Chakraborty. 1975. The Bottleneck Effect and Genetic Variability in Populations. *Evolution* 29 (1):1-10.
- Nordborg, M. 2007. Coalescent Theory. In *Handbook of Statistical Genetics*, edited by B. M. Balding D, Cannings C. New-York: John Wiley & Sons, Ltd.
- Norris, D. E., A. C. Shurtleff, Y. T. Touré, and G. C. Lanzaro. 2001. Microsatellite DNA Polymorphism and Heterozygosity Among Field and Laboratory Populations of *Anopheles gambiae* s.s. (Diptera: Culicidae). *Journal of Medical Entomology* 38 (2):336-340.
- Nowak, C., C. Vogt, J. B. Diogo, and K. Schwenk. 2007. Genetic impoverishment in laboratory cultures of the test organism *Chironomus riparius*. *Environmental Toxicology and Chemistry* 26 (5):1018-1022.
- Oh, W. S., P. Y. Jeong, H. J. Joo, J. E. Lee, Y. S. Moon, H. M. Cheon, J. H. Kim, Y. U. Lee, Y. H. Shim, and Y. K. Paik. 2009. Identification and Characterization of a Dual-Acting Antinematodal Agent against the Pinewood Nematode, *Bursaphelenchus xylophilus*. *PLoS One* 4 (11):A22-A30.
- Raymond, M. and F. Rousset. 1995. An exact test for population differentiation. *Evolution* 49 (6):1280-1283.
- Reed, David H. and Richard Frankham. 2003. Correlation between Fitness and Genetic Diversity. *Conservation Biology* 17 (1):230-237.
- Rousset, F. 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8 (1):103-106.
- Saitou, N and M Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4 (4):406-425.
- Simões, P., M. Pascual, M. M. Coelho, and M. Matos. 2010. Divergent evolution of molecular markers during laboratory adaptation in *Drosophila subobscura*. *Genetica* 138 (9-10):999-1009.
- Sokal, R. R. and F. J. Rohlf. 1995. *Biometry. The principles and practice of statistics in biological research*. Third Edition ed.
- Son, J., M. Komatsu, N. Matsushita, and T. Hogetsu. 2010. Migration of pine wood nematodes in the tissues of *Pinus thunbergii*. *Journal of Forest Research* 15 (3):186-193.
- Stohler, R. A., J. Curtis, and D. J. Minchella. 2004. A comparison of microsatellite polymorphism and heterozygosity among field and laboratory populations of *Schistosoma mansoni*. *International Journal for Parasitology* 34 (5):595-601.
- Stuart, R. J. and R. Gaugler. 1996. Genetic adaptation and founder effect in laboratory populations of the entomopathogenic nematode *Steinernema glaseri*. *Canadian Journal of Zoology* 74 (1):164-170.
- Tares, S., P. Abad, N. Bruguier, and G. de Guiran. 1992. Identification and evidence for relationships among geographical isolates of *Bursaphelenchus* spp. (pinewood nematode) using homologous DNA probes. *Heredity* 68 (2):157-164.
- R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

- Valadas, V., P. Barbosa, M. Espada, S. Oliveira, and M. Mota. 2012. The pinewood nematode, *Bursaphelenchus xylophilus*, in Portugal : possible introductions and spread routes of a serious biological invasion revealed by molecular methods. *Nematology* 00 (0):1 - 13.
- Wang, Y., T. Yamada, D. Sakaue, and K. Suzuki. 2005. Variations in life history parameters and their influence on rate of population increase of different pathogenic isolates of the pine wood nematode, *Bursaphelenchus xylophilus*. *Nematology* 7 (3):459-467.
- Weir, B. S. and C. Clark Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38 (6):1358-1370.
- Willi, Y., J. Van Buskirk, and A. A. Hoffmann. 2006. Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics* 37:433-458.
- Woodworth, L. M., M. E. Montgomery, D. A. Briscoe, and R. Frankham. 2002. Rapid genetic deterioration in captive populations: Causes and conservation implications. *Conservation Genetics* 3 (3):277-288.
- Wright, S. 1931. Evolution in mendelian populations. *Genetics* 16:97 - 159.
- Zhu, L., J. Ye, S. Negi, X. Xu, Z. Wang, and J. Ji. 2012. Pathogenicity of Aseptic *Bursaphelenchus xylophilus*. *PLoS One* 7 (5):e38095.

Table 1: Characteristics of the field samples and laboratory strains of PWN used in this study.

Code	Category	Duration in collection	Origin	No. of individuals	Mean Na	Max Na	He
MO1	Field	-	USA - Missouri	31	3,25	8	0,33
MO2	Field	-	USA - Missouri	23	3,38	7	0,36
NE1	Field	-	USA - Nebraska	16	2,00	4	0,20
NE2	Field	-	USA - Nebraska	15	1,81	4	0,21
NE5	Field	-	USA - Nebraska	14	1,94	5	0,21
NE6	Field	-	USA - Nebraska	21	1,38	2	0,18
NE9	Field	-	USA - Nebraska	29	2,63	5	0,31
NE10	Field	-	USA - Nebraska	26	1,75	4	0,19
NE12	Field	-	USA - Nebraska	28	1,50	3	0,15
NE13b	Field	-	USA - Nebraska	19	2,38	5	0,28
NE14	Field	-	USA - Nebraska	28	1,69	3	0,24
NE15	Field	-	USA - Nebraska	23	2,25	5	0,22
NE19	Field	-	USA - Nebraska	16	1,81	3	0,23
NE22	Field	-	USA - Nebraska	17	2,00	5	0,28
NE23	Field	-	USA - Nebraska	25	1,25	2	0,07
NE24	Field	-	USA - Nebraska	19	1,31	2	0,07
VI9	Field	-	USA - Virginia	22	1,50	3	0,16
VI10	Field	-	USA - Virginia	19	2,69	7	0,30
MA1	Field	-	USA - Massachusetts	19	1,50	2	0,14
NC1	Field	-	USA - North Carolina	12	2,81	6	0,38
NC2	Field	-	USA - North Carolina	23	3,13	7	0,36
NY1	Field	-	USA - New York	10	2,31	5	0,36
GA1	Field	-	USA - Georgia	7	1,94	4	0,26
GA2	Field	-	USA - Georgia	20	2,63	6	0,34
GA3	Field	-	USA - Georgia	8	2,75	6	0,40
GA4	Field	-	USA - Georgia	15	2,69	5	0,40
KS1	Field	-	USA - Kansas	17	3,94	10	0,46
Boylston	Field	-	USA - Mississippi	32	2,19	4	0,35
US2	Old	30	USA - Vermont	30	1,94	4	0,22
US9	Old	26	USA - Arizona	25	1	1	-
US10	Old	26	USA - Minnesota	26	1,88	3	0,19
USA345	Recent	5	USA	24	2,38	5	0,28
USA618	Recent	5	USA	25	2,50	5	0,34
USA745	Recent	5	USA	27	2,19	5	0,26
USDE5	Recent	2	USA	27	3,69	7	0,35
Alberta	Old	21	Canada - Alberta	21	1,50	3	0,14
BC	Old	24	Canada - British Columbia	30	1,44	4	0,15
Q52A	Old	24	Canada - Quebec	28	1,31	2	0,10
St John	Old	24	Canada - New Brunswick	23	1,63	3	0,13
St William	Old	24	Canada - Ontario	30	1,44	4	0,13

Note: The duration in collection is given in year and is not applicable for field samples. Mean Na is the mean number of alleles per sample, Max Na is the maximum number of alleles detected per locus in each sample and Mean He is the mean expected heterozygosity per sample.

FIGURE LEGENDS

Figure 1. Scenario of the collection maintenance of PWN, i.e., transplanting protocol, simulated during 15 years with the DIYABC software version 2.0 (Cornuet et al. 2014). N_i : effective population size in the field; N_p : effective population size in the inoculation periods and N_g : effective population size in the population growth periods. Each year (i.e., after two transplanting events), a sampling was performed to assess the genetic diversity of the simulated sample along the collection maintenance.

Figure 2. Genetic diversity per locus within each category of PWN: “Field”, “Recent” and “Old”. (A) Mean number of alleles (\pm SE) and (B) mean expected heterozygosity (\pm SE). “Field” corresponded to field samples, “Recent” to laboratory strains aged 5 years or less and “Old” to laboratory strains aged more than 20 years

Figure 3. Differentiation per locus between each pair of categories of PWN: “Field-Recent”, “Field-Old” and “Recent-Old”. (A) Corrected estimate of F_{ST} (\pm SE) and (B) D_{est} (\pm SE). “Field” corresponded to field samples, “Recent” to laboratory strains aged 5 years or less and “Old” to laboratory strains aged more than 20 years.

Figure 4. Differentiation per locus within each category of PWN: “Field”, “Recent” and “Old”. (A) Corrected estimate of F_{ST} (\pm SE) and (B) D_{est} (\pm SE). “Field” corresponded to field samples, “Recent” to laboratory strains aged 5 years or less and “Old” to laboratory strains aged more than 20 years.

Figure 5. Genetic diversity as a function of time spent in collection, obtained from the simulations of the maintenance protocol of PWN. (A) Mean number of alleles and (B) mean expected heterozygosity, both computed over 1,000 observations. N_p : effective size in the inoculation periods and N_g : effective size in the population growth periods. Uniform distributions were used with different ranges of values for both N_p and N_g , as shown in the legend (for more details see Material and Methods). The impact of an accident (a drastic reduction in effective population size after 270 generations in collection) on the genetic diversity was also shown.

SUPPORTING INFORMATION

Table S1. Differentiation between each pair of field samples and laboratory strains of PWN. Below the diagonal, pairwise Weir & Cockerham estimates of F_{ST} (Weir & Cockerham 1984), corrected for null alleles, obtained with FREENA (Chapuis & Estoup 2007). Above the diagonal, D_{est} of Jost (2008) obtained with SMOGD (Crawford 2010).

Figure S1. NJ tree with the three categories of PWN samples/strains: “Field”, “Recent” and “Old”. (A) Distances of Cavalli-Sforza & Edwards (Cavalli-Sforza & Edwards 1967), with bootstrap (on locus) values calculated over 2,000 replications given as percentages (only values > 20% are shown). (B) Distances of Cavalli-Sforza & Edwards (Cavalli-Sforza & Edwards 1967) corrected for null alleles, obtained with FREENA (Chapuis & Estoup 2007). The “Recent” strains and the “Old” strains were highlighted in green and orange, respectively.

Figure 1

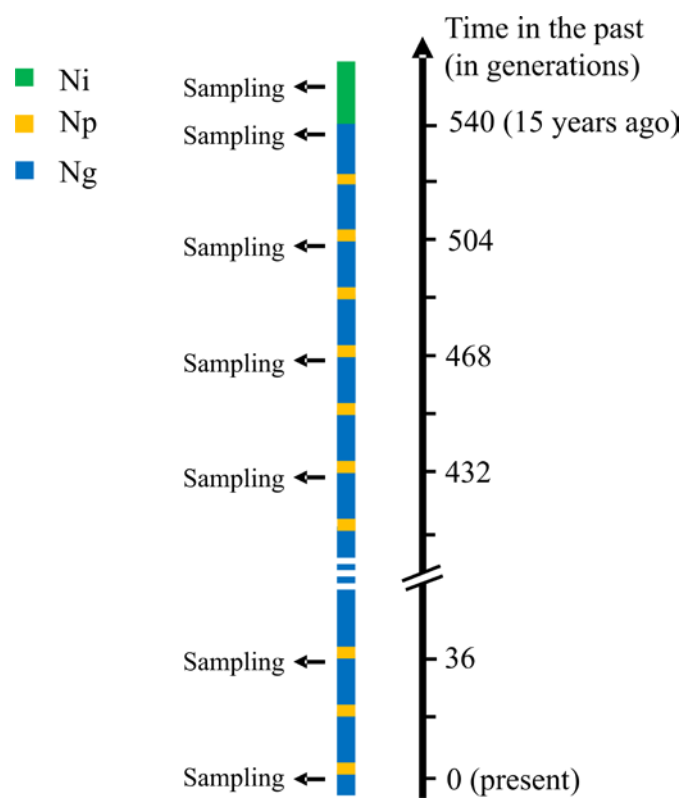


Figure 2

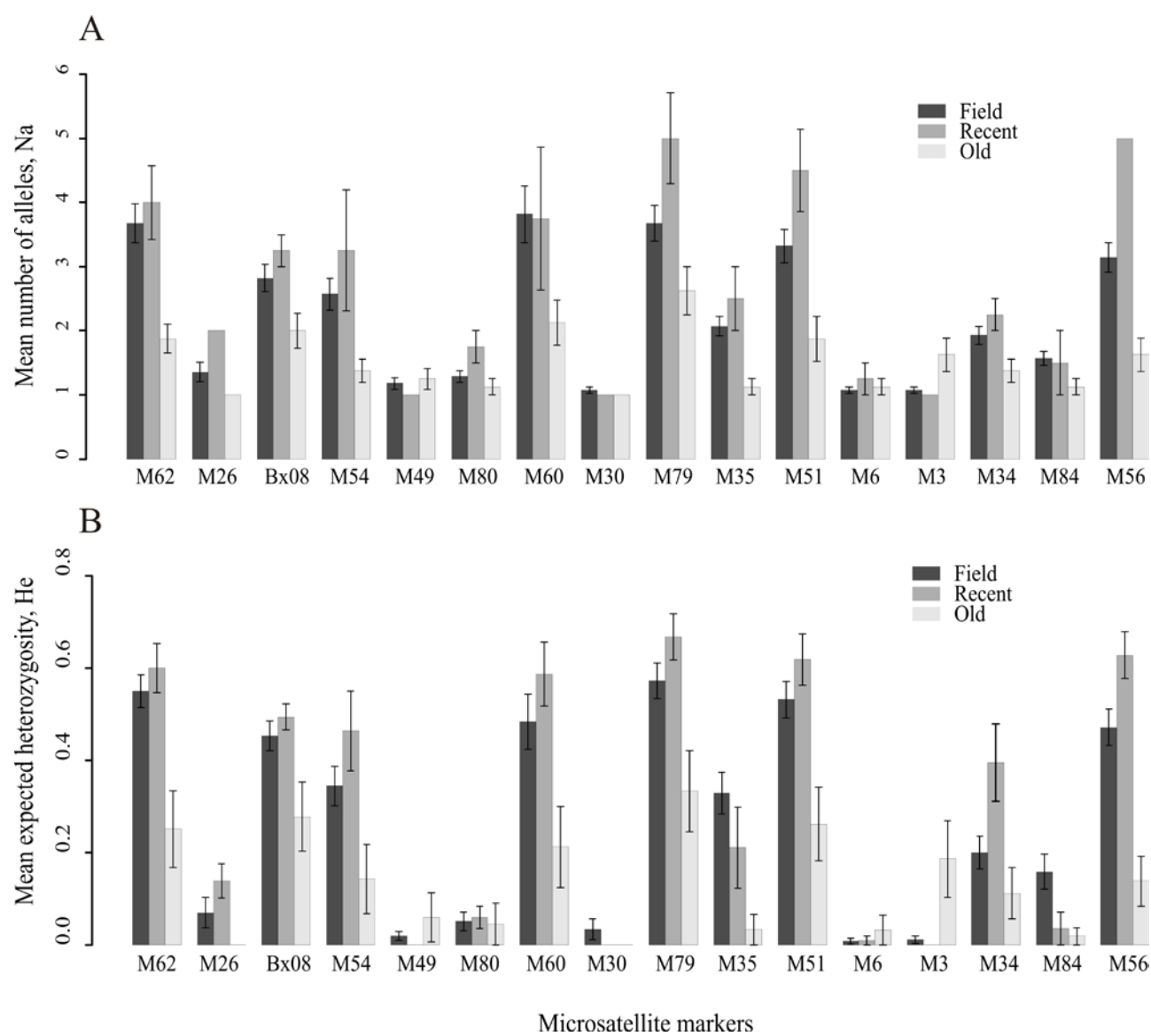


Figure 3

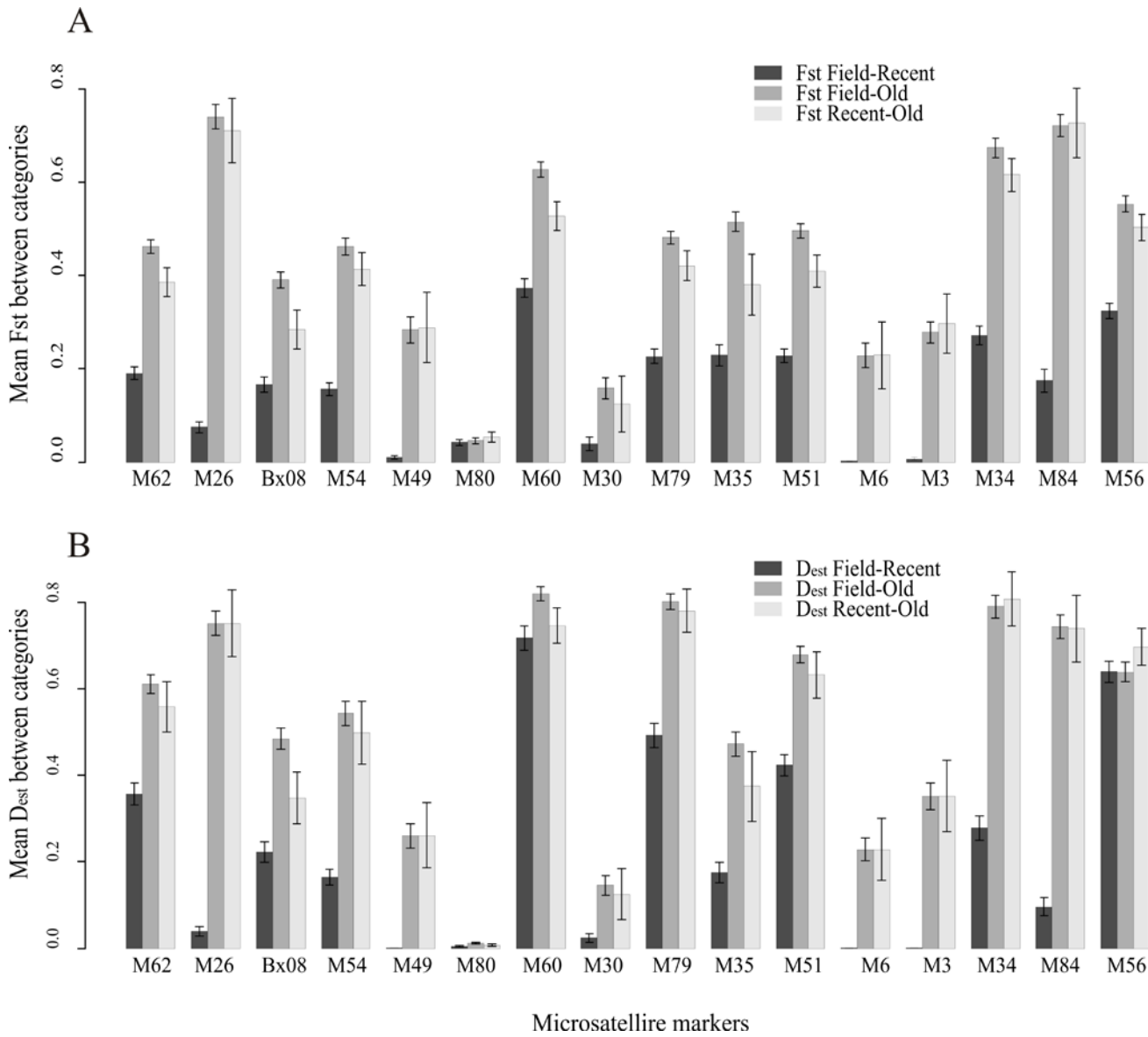


Figure 4

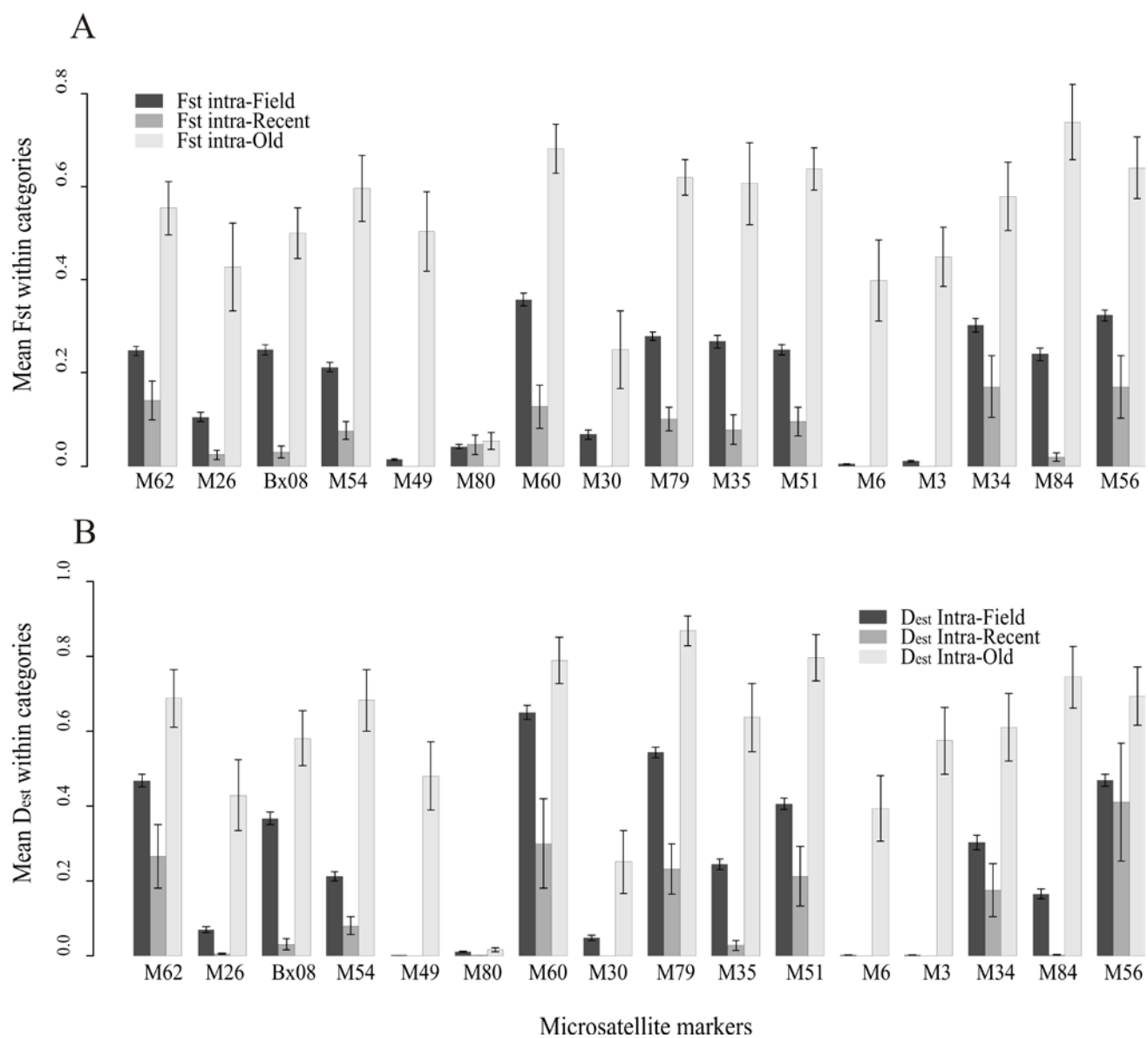


Figure 5

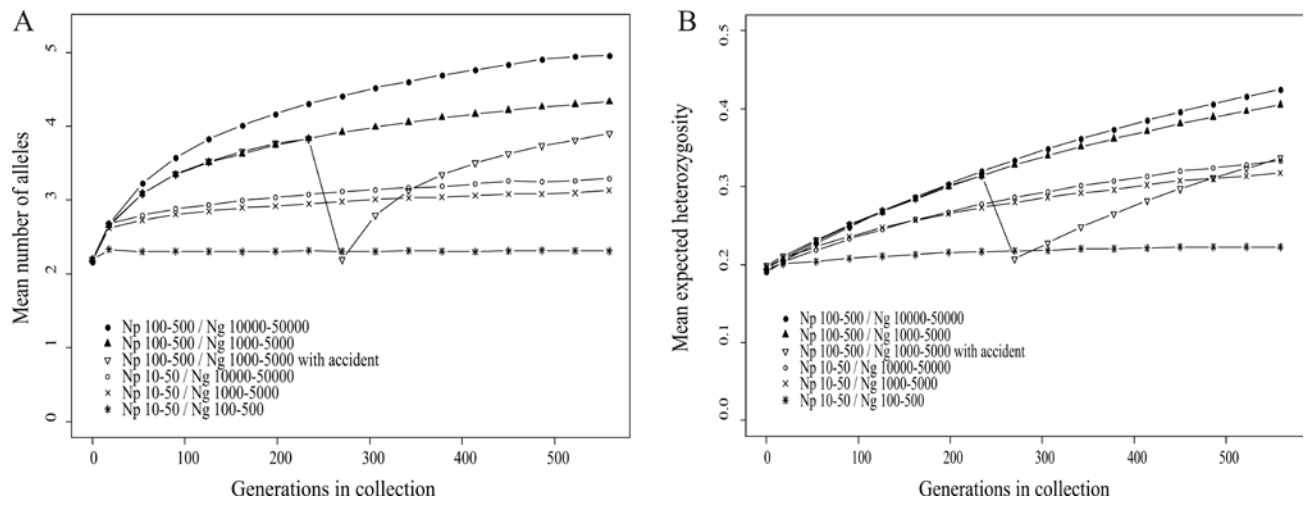
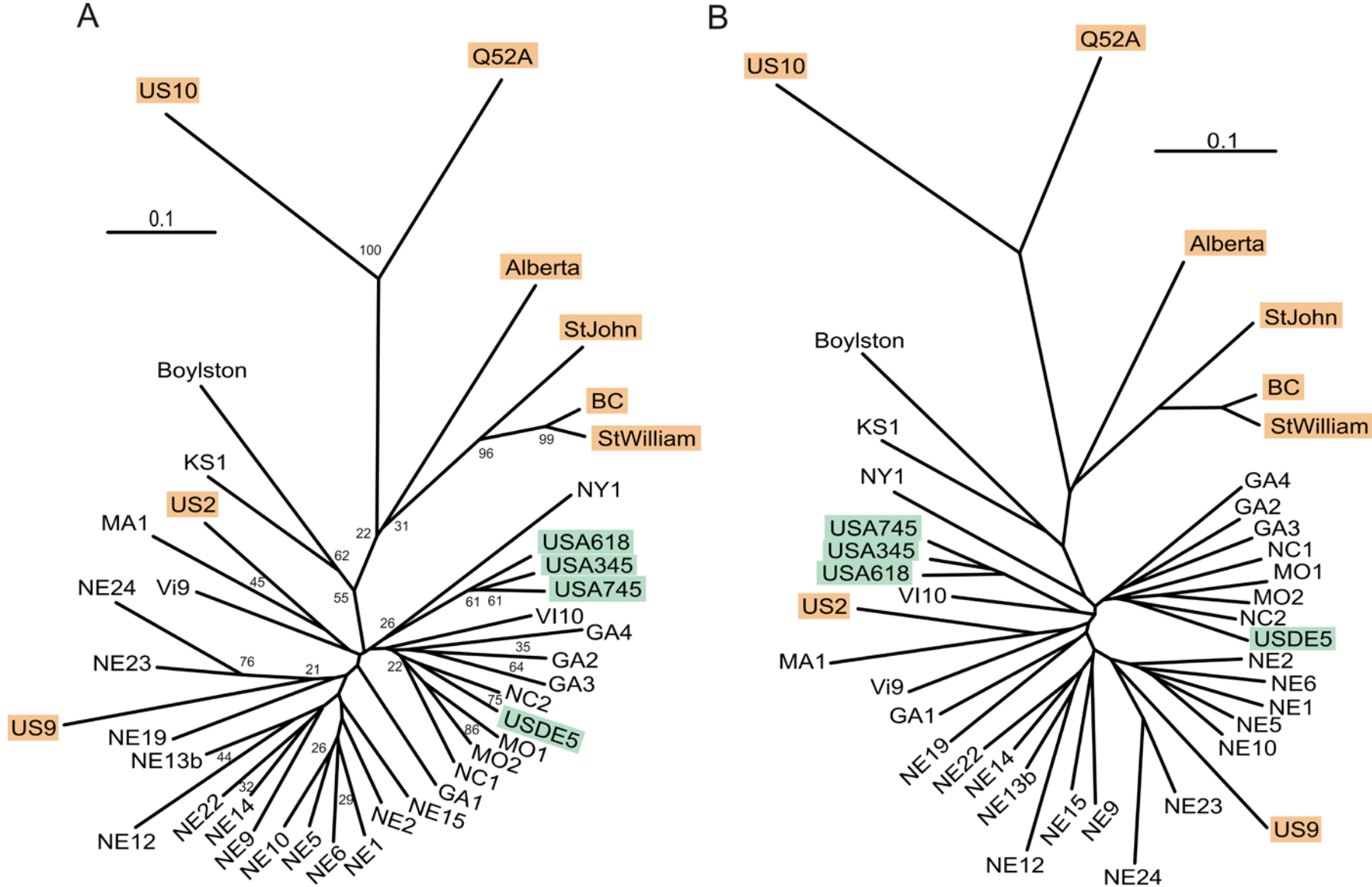


Table S1

	MO1	MO2	NE1	NE2	NE5	NE6	NE9	NE10	NE12	NE13b	NE14	NE15	NE19	NE22	NE23	NE24	Vi9	Vi10	MA1	NC1	NC2	NY1	GA1	GA2	GA3	GA4	KS1	Boylston	US2	US9	US10	USA345	USA618	USA745	USDE5	Alberta	BC	Q52A	StJohn	StWilliam	
MO1		0.020	0.085	0.059	0.088	0.059	0.062	0.102	0.095	0.032	0.060	0.099	0.060	0.056	0.106	0.150	0.096	0.048	0.161	0.056	0.054	0.136	0.080	0.057	0.044	0.048	0.080	0.197	0.074	0.153	0.657	0.058	0.053	0.084	0.034	0.313	0.244	0.499	0.309	0.264	
MO2	0.065		0.078	0.047	0.069	0.077	0.057	0.084	0.124	0.053	0.056	0.077	0.031	0.063	0.094	0.144	0.083	0.048	0.128	0.031	0.017	0.095	0.042	0.049	0.020	0.043	0.060	0.209	0.080	0.121	0.664	0.063	0.064	0.081	0.015	0.269	0.215	0.535	0.298	0.241	
NE1	0.293	0.266			0.029	0.008	0.018	0.024	0.014	0.104	0.048	0.033	0.018	0.056	0.043	0.043	0.094	0.087	0.103	0.207	0.103	0.065	0.118	0.086	0.079	0.069	0.077	0.159	0.306	0.176	0.130	0.658	0.064	0.063	0.065	0.069	0.273	0.186	0.626	0.212	0.206
NE2	0.245	0.194	0.188		0.012	0.029	0.020	0.034	0.119	0.045	0.043	0.042	0.071	0.016	0.046	0.127	0.092	0.091	0.191	0.087	0.071	0.119	0.079	0.091	0.071	0.083	0.122	0.309	0.143	0.083	0.677	0.061	0.065	0.075	0.066	0.293	0.197	0.616	0.199	0.218	
NE5	0.278	0.240	0.120	0.145		0.017	0.013	0.004	0.093	0.040	0.010	0.024	0.051	0.018	0.029	0.071	0.077	0.084	0.186	0.117	0.074	0.134	0.069	0.081	0.075	0.083	0.159	0.333	0.167	0.099	0.666	0.069	0.067	0.076	0.074	0.353	0.176	0.641	0.234	0.204	
NE6	0.274	0.257	0.168	0.161	0.141		0.016	0.028	0.070	0.031	0.023	0.059	0.084	0.038	0.054	0.113	0.105	0.086	0.155	0.105	0.098	0.191	0.098	0.074	0.096	0.101	0.173	0.320	0.129	0.117	0.666	0.092	0.100	0.111	0.093	0.376	0.251	0.570	0.290	0.264	
NE9	0.226	0.194	0.140	0.108	0.087	0.137		0.020	0.080	0.026	0.021	0.035	0.079	0.030	0.071	0.136	0.108	0.090	0.202	0.091	0.085	0.169	0.103	0.098	0.096	0.077	0.139	0.325	0.128	0.119	0.635	0.100	0.099	0.109	0.070	0.416	0.231	0.597	0.277	0.254	
NE10	0.327	0.291	0.180	0.250	0.086	0.278	0.153		0.105	0.033	0.022	0.018	0.066	0.052	0.039	0.080	0.061	0.084	0.189	0.119	0.075	0.132	0.107	0.082	0.107	0.082	0.174	0.294	0.184	0.131	0.717	0.102	0.090	0.112	0.058	0.362	0.176	0.682	0.260	0.220	
NE12	0.387	0.402	0.493	0.500	0.475	0.502	0.411	0.521		0.043	0.089	0.140	0.094	0.104	0.116	0.192	0.251	0.150	0.175	0.190	0.176	0.223	0.101	0.178	0.171	0.169	0.215	0.272	0.103	0.200	0.697	0.144	0.155	0.122	0.156	0.436	0.377	0.542	0.438	0.400	
NE13b	0.175	0.172	0.221	0.235	0.188	0.239	0.116	0.186	0.287		0.018	0.047	0.081	0.030	0.084	0.127	0.107	0.062	0.174	0.060	0.057	0.144	0.112	0.086	0.067	0.050	0.122	0.262	0.086	0.168	0.672	0.082	0.065	0.120	0.048	0.386	0.262	0.554	0.343	0.302	
NE14	0.220	0.196	0.227	0.188	0.154	0.153	0.108	0.212	0.424	0.118		0.040	0.083	0.025	0.053	0.104	0.042	0.056	0.152	0.061	0.054	0.125	0.078	0.077	0.075	0.064	0.120	0.306	0.079	0.111	0.619	0.073	0.076	0.087	0.052	0.367	0.219	0.495	0.277	0.242	
NE15	0.311	0.263	0.235	0.245	0.195	0.365	0.166	0.150	0.595	0.225	0.273		0.052	0.063	0.077	0.117	0.075	0.069	0.166	0.081	0.061	0.098	0.100	0.080	0.078	0.088	0.164	0.313	0.164	0.170	0.682	0.072	0.063	0.081	0.054	0.322	0.133	0.656	0.223	0.173	
NE19	0.192	0.150	0.319	0.311	0.292	0.381	0.290	0.341	0.439	0.242	0.322	0.353		0.064	0.060	0.091	0.108	0.098	0.147	0.099	0.044	0.147	0.045	0.067	0.045	0.080	0.153	0.222	0.128	0.128	0.720	0.069	0.080	0.077	0.046	0.233	0.207	0.551	0.285	0.219	
NE22	0.249	0.214	0.216	0.110	0.175	0.197	0.121	0.279	0.401	0.187	0.135	0.291	0.306		0.080	0.156	0.100	0.090	0.186	0.063	0.056	0.126	0.087	0.091	0.075	0.063	0.148	0.319	0.133	0.114	0.632	0.069	0.063	0.081	0.074	0.296	0.250	0.557	0.257	0.264	
NE23	0.413	0.390	0.505	0.447	0.411	0.503	0.430	0.422	0.662	0.469	0.416	0.569	0.480	0.476		0.021	0.085	0.096	0.195	0.134	0.089	0.166	0.108	0.084	0.089	0.110	0.205	0.242	0.232	0.094	0.864	0.097	0.096	0.114	0.072	0.266	0.263	0.729	0.261	0.273	
NE24	0.498	0.504	0.623	0.641	0.567	0.632	0.500	0.547	0.695	0.554	0.543	0.608	0.573	0.611	0.594		0.133	0.157	0.267	0.176	0.128	0.215	0.156	0.099	0.101	0.143	0.237	0.298	0.305	0.187	0.875	0.169	0.163	0.201	0.113	0.295	0.259	0.739	0.299	0.276	
Vi9	0.333	0.307	0.508	0.474	0.483	0.505	0.398	0.462	0.688	0.448	0.347	0.452	0.491	0.441	0.576	0.711		0.064	0.136	0.070	0.048	0.122	0.082	0.079	0.080	0.073	0.174	0.270	0.136	0.138	0.607	0.083	0.091	0.107	0.050	0.311	0.198	0.574	0.236	0.219	
Vi10	0.132	0.109	0.333	0.301	0.287	0.328	0.251	0.321	0.456	0.219	0.216	0.266	0.283	0.285	0.474	0.570	0.317		0.087	0.037	0.041	0.096	0.081	0.054	0.036	0.054	0.111	0.231	0.088	0.180	0.715	0.058	0.049	0.074	0.028	0.343	0.217	0.581	0.295	0.266	
MA1	0.427	0.387	0.596	0.553	0.582	0.567	0.482	0.609	0.611	0.502	0.487	0.598	0.541	0.510	0.715	0.759	0.604	0.378		0.131	0.126	0.182	0.148	0.131	0.160	0.186	0.229	0.231	0.116	0.277	0.760	0.099	0.109	0.120	0.133	0.353	0.301	0.660	0.298	0.328	
NC1	0.126	0.074	0.305	0.238	0.309	0.317	0.206	0.363	0.475	0.215	0.213	0.281	0.270	0.217	0.472	0.524	0.325	0.118	0.405		0.010	0.078	0.097	0.024	0.019	0.066	0.080	0.261	0.095	0.194	0.679	0.077	0.076	0.100	0.026	0.275	0.212	0.583	0.251	0.255	
NC2	0.112	0.055	0.253	0.221	0.243	0.280	0.215	0.277	0.427	0.194	0.185	0.225	0.167	0.205	0.391	0.474	0.257	0.094	0.379	0.057		0.070	0.045	0.035	0.003	0.028	0.064	0.222	0.096	0.165	0.684	0.052	0.053	0.069	0.002	0.233	0.168	0.583	0.242	0.205	
NY1	0.258	0.217	0.367	0.355	0.342	0.432	0.317	0.376	0.525	0.301	0.323	0.319	0.310	0.311	0.580	0.589	0.418	0.236	0.474	0.198	0.157		0.083	0.109	0.054	0.098	0.134	0.262	0.223	0.215	0.721	0.073	0.057	0.053	0.064	0.275	0.187	0.676	0.224	0.242	
GA1	0.207	0.145	0.387	0.330	0.323	0.397	0.296	0.407	0.521	0.313	0.280	0.377	0.225	0.297	0.579	0.668	0.449	0.229	0.540	0.224	0.129	0.248		0.094	0.050	0.083	0.111	0.258	0.099	0.144	0.604	0.060	0.057	0.045	0.058	0.327	0.158	0.463	0.284	0.182	
GA2	0.136	0.101	0.265	0.257	0.254	0.283	0.232	0.298	0.460	0.239	0.255	0.265	0.217	0.266	0.387	0.429	0.334	0.157	0.414	0.106	0.092	0.242	0.230		0.010	0.050	0.109	0.237	0.145	0.155	0.748	0.106	0.096	0.118	0.021	0.285	0.161	0.663	0.216	0.187	
GA3	0.112	0.084	0.279	0.269	0.281	0.333	0.247	0.340	0.458	0.217	0.245	0.285	0.182	0.252	0.504	0.507	0.356	0.123	0.441	0.072	0.028	0.131	0.135	0.053		0.010	0.049	0.205	0.127	0.182	0.675	0.044	0.039	0.059	0.005	0.233	0.156	0.566	0.209	0.191	
GA4	0.085	0.092	0.263	0.257	0.256	0.300	0.202	0.291	0.435	0.176	0.205	0.254	0.203	0.230	0.432	0.486	0.308	0.120	0.441	0.113	0.064	0.165	0.185	0.094	0.045		0.072	0.201	0.134	0.156	0.597	0.074	0.071	0.099	0.022	0.302	0.181	0.550	0.247	0.203	
KS1	0.236	0.217	0.357	0.314	0.332	0.385	0.295	0.384	0.448	0.287	0.302	0.355	0.299	0.284	0.502	0.510	0.424	0.252	0.430	0.207	0.192	0.211	0.208	0.228	0.133	0.183		0.143	0.180	0.269	0.652	0.137	0.103	0.124	0.052	0.273	0.150	0.506	0.177	0.168	
Boylston	0.338	0.334	0.477	0.452	0.464	0.497	0.426	0.493	0.492	0.403	0.448	0.504	0.396	0.428	0.537	0.558	0.520	0.369	0.467	0.350	0.323	0.360	0.395	0.344	0.294	0.301	0.233		0.303	0.372	0.752	0.241	0.237	0.253	0.205	0.289	0.284	0.527	0.266	0.275	
US2	0.246	0.238	0.482	0.443	0.506	0.471	0.364	0.541	0.522	0.345	0.373	0.516	0.447	0.422	0.685	0.719	0.548	0.284	0.495	0.266	0.289	0.453	0.421	0.341	0.351	0.310	0.367	0.477		0.201	0.650	0.131	0.132	0.150	0.105	0.519	0.358	0.416	0.442	0.390	
US9	0.602	0.580	0.798	0.730	0.767	0.765	0.594	0.744	0.825	0.715	0.651	0.770	0.736	0.686	0.854	0.929	0.821	0.70																							

Figure S1



3. Routes d'invasion par la méthode ABC

Cette partie, pas encore pleinement aboutie, présente les analyses que nous avons effectuées avec la méthode ABC pour tenter de clarifier les routes d'invasion du PWN. L'étude a été réalisée en collaboration avec Éric Lombaert.

❖ Introduction

La plupart des méthodes d'inférence basées sur des modèles impliquent le calcul de la fonction de vraisemblance, qui est centrale mais difficile à obtenir lorsque les modèles sont trop complexes. L'ABC contourne le calcul de la vraisemblance *via* des simulations intensives et l'utilisation de statistiques résumées, et permet ainsi de traiter les modèles complexes (Bertorelle et al. 2010; Csillery et al. 2010; Sunnaker et al. 2013) que l'on peut rencontrer lorsque l'on étudie l'histoire d'introduction d'organismes envahissants. Cette méthode permet aussi de combiner des données génétiques (issues de marqueurs microsatellites par exemple), des données historiques (dates de première observation) et des observations de terrain/données biologiques grâce à son cadre Bayésien (Beaumont & Rannala 2004). De plus, elle prend en compte des statistiques résumées (i.e., statistiques résumant la variation génétique existante entre et au sein des populations étudiées) simplifiant la complexité des jeux de données génétiques. Cependant, elle en utilise plusieurs, augmentant le niveau d'informations utilisées pour les inférences par rapport aux méthodes traditionnelles (arbres fondés sur les distances génétiques, vraisemblances d'assignation).

Dans le cas du PWN, les méthodes indirectes traditionnelles ne nous ont pas permis d'identifier l'origine des populations européennes (voir étude présentée partie III.1.b). L'ABC présente plusieurs avantages par rapport aux méthodes traditionnelles dans ce contexte,

comme décrits dans la partie I.3.b. Nous avons donc utilisé cette méthode pour tenter de clarifier les routes d'invasion du PWN et déterminer le scénario d'invasion le plus probable. Avant de présenter l'étude que nous avons réalisée, nous pensons utile d'expliquer concrètement ce qu'est une analyse ABC.

Une analyse ABC se déroule en 4 grandes étapes. (i) La première étape consiste à définir les modèles, les distributions de densité *a priori* (aussi appelées *priors*) des différents paramètres et les probabilités *a priori* des modèles. Les modèles correspondent aux différents scénarii d'invasion que l'on souhaite comparer (introductions unique vs. multiples, indépendantes vs. successives,... voir le matériel et méthodes pour des exemples). Les paramètres, quant à eux, sont d'ordre démographique (par ex. les tailles efficaces), historique (par ex., les dates de première observation) et mutationnel (par ex., le taux de mutation des marqueurs utilisés et le modèle mutationnel). (ii) La seconde étape est l'étape de simulation. On simule, en tirant les valeurs des paramètres dans les *priors*, un grand nombre de jeux de données sous chaque scénario grâce au modèle de coalescence (Nordborg 2007). On choisit les statistiques résumées que l'on souhaite utiliser (classiquement, le nombre d'allèles, l'hétérozygotie attendue, une mesure de différenciation entre populations...). Chaque jeu de données sera caractérisé par les statistiques résumées choisies. (iii) Vient ensuite l'étape de comparaison de scénarii et de sélection du scénario le plus probable. La comparaison des différents scénarii se fait en confrontant les statistiques résumées observées (celles issues des données obtenues par analyse des populations naturelles échantillonnées) et simulées. Les jeux de données simulées dont les statistiques résumées sont les plus distantes des statistiques résumées observées sur la base de distances euclidiennes sont rejetés. Les jeux de données simulées les plus proches du jeu de données observées permettent ensuite de calculer des probabilités *a posteriori* et leur intervalle de confiance par une régression logistique multinomiale. Ceci permet de sélectionner le scénario d'invasion le plus probable, celui ayant

la probabilité *a posteriori* la plus élevée. Les paramètres du scénario choisi peuvent alors être estimés par une méthode similaire. Pour chaque paramètre, une régression linéaire permet d'obtenir sa distribution *a posteriori*. (iv) Nous devons ensuite évaluer la confiance que l'on peut avoir dans le choix du scénario sélectionné. Cette étape de vérification constitue la quatrième et dernière étape. Elle se fait *via* différents moyens. On vérifie tout d'abord le non-recoupement des intervalles de confiance associés à chaque probabilité *a posteriori*. On calcule ensuite les erreurs de type I (i.e., la probabilité de ne pas sélectionner un scénario lorsqu'il est "vrai") et de type II (i.e., la probabilité de sélectionner le scénario alors qu'il est "faux") sur la base de données simulées selon les scénarii considérés. Ce calcul d'erreurs permet de s'assurer que les différents scénarii comparés sont bien différenciables. On vérifie enfin l'adéquation du scénario sélectionné aux données observées par de nouvelles simulations. Les valeurs des paramètres, utilisées pour les simulations, sont cette fois-ci tirées dans les distributions *a posteriori* des paramètres du scénario sélectionné. Cela nous permet d'évaluer si le scénario sélectionné et ses *posteriors* produisent des données qui sont proches des données observées.

❖ Matériel et méthodes

Cycle de vie et temps de génération

Comme évoqué dans l'introduction (partie II.1), le cycle biologique du PWN *in natura* est complexe (Evans et al. 1996; Futai 2013 pour une revue). Le PWN et son vecteur sont étroitement associés et le PWN s'est adapté à son vecteur (Mamiya 1972). Le PWN se reproduit donc exclusivement en été une fois qu'il a envahi un arbre sensible. Son temps de génération *in natura* n'est pas précisément connu en raison de l'impossibilité d'observer directement le PWN dans l'arbre. Néanmoins, nous savons que la durée du cycle de vie du

PWN dépend fortement de la température (Mamiya 1972; Mamiya 1975) : de 4-5 jours à 25°C à environ 12 jours à 15°C. Par conséquent, en couplant ces informations, nous avons déduit que le temps de générations du PWN était d'environ 30.5 générations par an, en considérant que le PWN se reproduit de Juin à Septembre et en estimant que la température moyenne en été est de 25°C dans son aire de distribution.

Échantillons analysés et génotypage

Nous avons analysés 48 échantillons en provenance de l'aire native et des différentes zones envahies : 28 échantillons des USA (554 individus) pour l'aire native, 7 échantillons du Japon (210 individus), 9 du Portugal (169 individus) et 4 de Chine (147 individus) pour les zones envahies. Les caractéristiques des échantillons sont présentées dans le tableau 1. Tous les individus ont été extraits à partir d'échantillons de bois collectés directement sur le terrain, grâce à un tamis ou la méthode de Baermann (Viglierchio & Schmitt 1983). Les connaissances historiques de l'invasion du PWN et celle de l'expansion des populations au sein de chaque zone envahie ont permis d'associer des dates d'occurrence aux différents échantillons, présenté dans le tableau 1.

Nous avons extrait l'ADN de chaque individu par choc thermique (Castagnone et al. 2005) et nous avons amplifié les 16 marqueurs microsatellites développés précédemment comme décrit par Mallez et al. (2014). Le marqueur M26 n'a pas amplifié l'un des échantillons chinois. Il a donc été exclu et les analyses présentées ci-après ont été finalement réalisées avec 15 marqueurs microsatellites.

Tableau 1 : Caractéristiques des échantillons analysés dans cette étude.

Code	Nb d'individus	Origine géographique	Année d'introduction / observation	Na moyen	Max Na	He	Ho	Fis
MO1	31	USA - Missouri - Columbia	-	3,33	8	0,35	0,29	0.19*
MO2	23	USA - Missouri - Columbia	-	3,53	7	0,38	0,28	0.27*
NE1	16	USA - Nebraska - Davey	-	2,07	4	0,21	0,19	0,09
NE2	15	USA - Nebraska - Davey	-	1,87	4	0,22	0,17	0,24
NE5	14	USA - Nebraska - Pawnee Lake	-	2,00	5	0,22	0,19	0,14
NE6	21	USA - Nebraska - Pawnee Lake	-	1,40	2	0,20	0,10	0.48*
NE9	29	USA - Nebraska - Pawnee Lake	-	2,73	5	0,33	0,25	0.26*
NE10	26	USA - Nebraska - Pawnee Lake	-	1,80	4	0,20	0,16	0.19*
NE12	28	USA - Nebraska - Conestoga Lake	-	1,53	3	0,16	0,11	0.34*
NE13b	19	USA - Nebraska - Pioneers Park	-	2,40	5	0,30	0,21	0.31*
NE14	28	USA - Nebraska - Pioneers Park	-	1,73	3	0,25	0,17	0.34*
NE15	23	USA - Nebraska - Pioneers Park	-	2,27	5	0,23	0,14	0.39*
NE19	16	USA - Nebraska - UNL East Campus	-	1,87	3	0,25	0,21	0,15
NE22	17	USA - Nebraska - Lincoln	-	2,07	5	0,30	0,22	0.28*
NE23	25	USA - Nebraska - Lincoln	-	1,27	2	0,08	0,07	0,09
NE24	19	USA - Nebraska - Lincoln	-	1,33	2	0,08	0,07	0,15
VI9	22	USA - Virginia - Midlothian	-	1,53	3	0,15	0,16	0,14
VI10	19	USA - Virginia - Virginia Beach	-	2,80	7	0,32	0,26	0,19
MA1	19	USA - Massachusetts - Worchester	-	1,53	2	0,17	0,15	0,00
NC1	12	USA - North Carolina	-	2,93	6	0,41	0,38	0,06
NC2	23	USA - North Carolina	-	3,27	7	0,38	0,30	0.22*
NY1	10	USA - New York	-	2,40	5	0,37	0,20	0.47*
GA1	7	USA - Georgia	-	2,00	4	0,27	0,23	0,16
GA2	20	USA - Georgia	-	2,67	6	0,36	0,31	0.13*
GA3	8	USA - Georgia	-	2,80	6	0,39	0,28	0,31
GA4	15	USA - Georgia	-	2,80	5	0,42	0,28	0.34*
KS1	17	USA - Kansas	-	3,87	10	0,44	0,36	0.18*
Boylston	32	USA - Mississippi - Boylston	-	2,20	4	0,34	0,24	0.29*
Jap120	23	Japan - Iwate - Shiwa	1978	1	1	-	-	-
Jap212	27	Japan - Iwate - Shiwa	1978	1	1	-	-	-
Jap308	25	Japan - Iwate - Shiwa	1978	1	1	-	-	-
Kasumig2	36	Japan - Ibaraki - Kasumigaura	1971	1	1	-	-	-
Kasumig3	29	Japan - Ibaraki - Kasumigaura	1971	1,20	2	0,07	0,07	-0,02
Kasumig5	35	Japan - Ibaraki - Kasumigaura	1971	1	1	-	-	-
Kosa	35	Japan - Kumamoto - Kosa	1905	1,27	2	0,06	0,06	0,03
Mad23PC	12	Madeira Island - Porto da Cruz	2008	1	1	-	-	-
Mad24C	7	Madeira Island - Calheta	2008	1	1	-	-	-
128S	17	Portugal - Setubal - Grândola	1999	1,07	2	0,03	0,01	0.62*
TR1	30	Portugal - Setubal - Troia	1999	1	1	-	-	-
TR2	27	Portugal - Setubal - Troia	1999	1	1	-	-	-
AM2	21	Portugal - Setubal - Aguas de Moura	1999	1	1	-	-	-
Comporta	28	Portugal - Setubal - Comporta	1999	1	1	-	-	-
E182	13	Portugal - Coimbra - Penela	2008	1	1	-	-	-
E1069	14	Portugal - Viseu - Castro Daire	2008	1	1	-	-	-
GuangDong	41	Chine - Guangdong	1988	1	1	-	-	-
GuangXi	32	Chine - GuangXi	-	1	1	-	-	-
Laoshan	39	Chine - Shandong - Laoshan	1982	1	1	-	-	-
ShanDong	35	Chine - Shandong	1982	1	1	-	-	-

Note : Na moyen, nombre moyen d'allèles par échantillon ; Max Na, nombre maximum d'allèles par locus dans chaque échantillon ; He, hétérozygotie attendue et Ho, hétérozygotie observée. *FIS* a été calculé comme décrit par Weir & Cockerham (1984). "*" indique que le résultat du test de HWE est significatif après correction. "-" indique que He, Ho et *FIS* ne sont pas calculés pour des échantillons avec seulement des marqueurs monomorphes et que la date d'introduction ne s'applique ou n'est pas connue.

Analyses génétiques classiques

Pour chaque échantillon, nous avons déterminé le nombre maximum d'allèles par locus (Max Na), le nombre moyen d'allèles (Na moyen) et les hétérozygoties observée (Ho) et attendue (He) par échantillon avec Genetix version 4.05 (Belkhir et al. 1996-2004). Nous avons testé si les échantillons étaient à l'équilibre de Hardy-Weinberg (HWE) avec Genepop version 4.1.3 (Rousset 2008) et nous avons quantifié les écarts à HWE inférés en calculant le Fis de Weir & Cockerham (1984) avec Fstat version 2.9.3.2 (Goudet 2002). Nous avons aussi testé la présence de déséquilibres de liaison (DL) entre marqueurs avec Genepop (Rousset 2008). Les tests multiples réalisées dans le cadre des tests de HWE et de DL ont été pris en compte en réalisant une correction FDR (pour False Discovery Rate, Benjamini & Hochberg 1995) et une correction de Bonferroni séquentiel (Sokal & Rohlf 1995), respectivement.

Relation entre les différents échantillons

Cette partie correspond aux analyses de génétique des populations faites à l'aide des méthodes traditionnelles présentées dans la partie I.3.b. Elles ont principalement été réalisées pour (i) déterminer le niveau de structure génétique rencontrée en Chine (que nous n'avons encore jamais étudié jusqu'ici), (ii) sélectionner les échantillons (particulièrement de l'aire native) qui seront inclus dans les analyses ABC et (iii) déterminer les scénarii d'invasion à comparer dans ces mêmes analyses.

Nous avons tout d'abord effectué des tests de différenciation génotypique entre paires d'échantillons au sein de chaque zone (native et envahie), entre zones envahies, et entre zones envahies et zone native, avec Genepop (Rousset 2008). Des tests multiples et non indépendants ayant été réalisés, le seuil de significativité a donc été ajusté par la méthode de Bonferroni séquentiel (Sokal & Rohlf 1995). Nous avons aussi calculé le Fst (Weir & Cockerham 1984), corrigés pour les allèles nuls (détectés pour certains échantillons et

marqueurs, voir Mallez et al. 2014), avec FreeNA (Chapuis & Estoup 2007) et le D_{est} (Jost 2008) avec SMOGD (Crawford 2010) entre paires d'échantillons.

Nous avons ensuite réalisé une analyse Bayésienne de groupement des individus avec le logiciel Structure version 2.3 (Pritchard et al. 2000) avec les échantillons des différentes zones native et envahies. Nous avons fait varier K (le nombre de clusters) de 1 à 10 et 20 séries indépendantes par valeur de K ont été réalisées, *via* la méthode de Monte Carlo par chaîne de Markov avec 10^6 itérations, précédées par une période de "rodage" de 2×10^5 itérations. Nous avons utilisé Clumpp (Jakobsson & Rosenberg 2007) pour identifier les patrons de groupement les plus fréquents parmi les 20 séries réalisées pour chaque valeur de K et nous avons représenté graphiquement ces derniers à l'aide de Distruct version 1.1 (Rosenberg 2004). L'analyse des représentations graphiques pour chaque valeur de K nous a permis de déterminer les relations existantes entre les différentes zones. Les relations entre les échantillons des différentes zones ont également été évaluées grâce à un arbre de distances génétiques basé sur les distances de Cavalli-Sforza & Edwards (Cavalli-Sforza & Edwards 1967), corrigées pour les allèles nuls, avec FreeNA (Chapuis & Estoup 2007). Nous avons finalement mesuré, grâce au logiciel GeneClass2 version 2.0 (Piry et al. 2004), la vraisemblance d'assignation des échantillons des zones envahies envers les échantillons américains.

La structure des populations en Chine a aussi été étudiée grâce à une analyse de groupement d'individus avec Structure version 2.3 (Pritchard et al. 2000), réalisée comme décrit précédemment avec K variant de 1 à 4. Le nombre de groupes a été déterminé par la méthode d'Evanno et al. (2005), automatisée dans Structure Harvester (http://taylor0.biology.ucla.edu/struct_harvest/index.php, Earl & vonHoldt 2012).

Analyses ABC

Toutes les analyses ABC ont été réalisées avec le logiciel DIYABC version 2.0.4 (Cornuet et al. 2014) dans la mesure où ses conditions d'application étaient remplies dans notre cas (reproduction sexuée et absence de flux de gènes entre populations, vérifiée par l'observation d'une très forte différenciation génétique significative entre quasiment tous les échantillons, voir Résultats).

Avant de réaliser des comparaisons de scénarii d'invasion à proprement parler, nous avons réalisé des analyses ABC préliminaires pour définir les *priors* démographiques et mutationnels les plus appropriés. Ces analyses préliminaires ont été réalisées avec des échantillons natifs. Un scénario simple a été défini : une population ancestrale native donne plusieurs échantillons natifs. Ce scénario était commun aux différents scénarii comparés, seules variaient la gamme des tailles efficaces et la présence ou non d'évènements de goulot d'étranglement. L'ajustement des *priors* s'est fait en vérifiant l'adéquation des données observées aux *priors* définis d'une part (option "evaluate scenarios and priors" du logiciel) et aux *posteriors* du scénario gagnant d'autre part (option "model checking" du logiciel). Ces analyses ont contribué à définir le groupe de *priors* n°1, qui a été utilisé comme base pour toutes les analyses ABC ultérieures (voir Tableau 2 pour la description des *priors*).

Le choix des scénarii à comparer est une étape cruciale dans l'ABC. Même quand le nombre d'échantillons est petit, beaucoup de scénarii sont possibles. Nous avons donc pris en compte l'information historique et les résultats de l'étude des relations entre échantillons pour définir les scénarii en compétition. Ainsi, les scénarii qui n'étaient pas supportés par les données génétiques ou historiques de manière évidente n'ont pas été considérés. Les résultats de l'étude des relations entre les différents échantillons ont mis en évidence que l'échantillon NE2 (échantillon natif) était globalement le plus proche des échantillons des zones envahies (voir Résultats). Il a donc été intégré dans les scénarii des analyses ABC pour représenter

l'aire native. Au vu de la forte structuration existante dans l'aire native, il est fort peu probable que NE2 soit la "vraie" source. Une population fantôme, partant de NE2, a donc également été incluse. Les résultats ont également montré que le Japon et la Chine présentent peut-être une histoire d'invasion plus complexe que nous aurions pu le penser au vu de la diversité génétique existante dans ces zones (voir Résultats). Nous avons donc procédé à des analyses ABC étape par étape et avons tenté de clarifier l'histoire asiatique avant d'intégrer l'Europe dans les scénarii.

Ainsi, nous avons testé dans un premier temps s'il y avait eu une ou plusieurs introductions en Chine, et plus précisément si Guangdong était ou non indépendant de Laoshan (voir Figure 4 pour la représentation des scénarii en compétition). Nous avons fait de même au Japon et nous nous sommes notamment concentrés sur l'échantillon Kasumig3. Nous avons essayé de déterminer s'il venait ou non de Kosa (l'échantillon le plus proche du point d'introduction au Japon, voir Figure 5 pour la représentation des différents scénarii en compétition). Nous avons ensuite étudié les relations entre le Japon et la Chine par plusieurs analyses ABC en tenant compte des résultats obtenus précédemment. Nous avons ainsi évalué si Laoshan venait du Japon ou des USA, de même pour Guangdong (voir Figures 6 et 7 pour la représentation des scénarii en compétition dans chaque analyse). Nous avons enfin intégré le Portugal dans le scénario d'invasion asiatique obtenu afin de déterminer son origine. Nous avons pour cela mis en compétition les 3 origines possibles pour le Portugal : les USA, le Japon ou la Chine (voir Figure 8 pour la représentation des scénarii en compétition).

Tableau 2 : Définition et distributions *a priori* des paramètres utilisées dans les scénarii d'invasion considérés dans les analyses ABC.

Description du paramètre	Nom du paramètre	Nb Générations/an	Distribution <i>a priori</i>	
			Forme	Intervalle
Paramètres démographiques			Groupe 1 / Groupe 2	Groupe 1 / Groupe 2
Taille efficace stable de la population ancestrale	Na		Log-Uniforme / Uniforme	[100 ; 10000]
Taille efficace stable des populations "fantôme"	Ng1 ; Ng2 ; Ng3		Log-Uniforme / Uniforme	[100 ; 10000]
Taille efficace stable de la population native, NE2	N1		Log-Uniforme / Uniforme	[100 ; 10000]
Taille efficace stable des populations envahissantes	NKo ; NKa ; NL ; NGu ; NT		Log-Uniforme / Uniforme	[100 ; 10000]
Taille efficace réduite pendant la fondation des populations envahissantes	NKoB ; NKaB ; NLB ; NGuB ; NTB		Log-Uniforme / Uniforme	[2 ; 100]
Temps de divergence* et fondation (en générations)				
Temps de divergence de NE2	t1		Uniforme / Log-Uniforme	[100 ; 10000]
Temps de divergence des populations fantôme	tg1 ; tg2 ; tg3 ; tadm		Uniforme / Log-Uniforme	[100 ; 10000]
Temps de divergence de Kosa	tKo	30	Uniforme / Log-Uniforme	[3294 ; 3873]
		15	Uniforme / Log-Uniforme	[1620 ; 1905]
		45	Uniforme / Log-Uniforme	[4860 ; 5715]
Temps de divergence de Kasumig3	tKa ; tKa1 ; tKa2 ; tKa3* ¹	30	Uniforme / Log-Uniforme	[1281 ; 1860]
Temps de divergence de Laoshan	tL ; tL1 ; tL2	30	Uniforme / Log-Uniforme	[945 ; 1525]
		15	Uniforme / Log-Uniforme	[465 ; 750]
		45	Uniforme / Log-Uniforme	[1395 ; 2250]
Temps de divergence de Guangdong	tGu ; tGu1 ; tGu2	30	Uniforme / Log-Uniforme	[762 ; 1342]
		15	Uniforme / Log-Uniforme	[375 ; 660]
		45	Uniforme / Log-Uniforme	[1125 ; 1980]
Temps de divergence de TR1	tT ; tT1 ; tT2 ; tT3	30	Uniforme / Log-Uniforme	[427 ; 1006]
		15	Uniforme / Log-Uniforme	[210 ; 495]
		45	Uniforme / Log-Uniforme	[630 ; 1485]
Durée de la fondation des populations	DB1 ; DB2 ; DB3 ; DB4		Uniforme / Log-Uniforme	[0 ; 100] / [1 ; 100]
Évènement d'admixture				
Taux d'admixture	ra		Uniforme	[0.1 ; 0.9]
Paramètres mutationnels				
Taux de mutation moyen	Mμ		Uniforme	[10 ⁻⁶ ; 10 ⁻⁴]
Paramètre moyen de la distribution géométrique du nombre de répétitions	MP		Uniforme	[10 ⁻⁸ ; 10 ⁻⁴]
Taux moyen d'insertion/délétion d'un seul nucléotide	Mμsni		Log-Uniforme	[0.1 ; 0.3]

* 20 ans d'incertitude ont été pris en compte pour les temps de divergence des populations envahissantes, ce qui correspond à l'intervalle des distributions présentées.

*¹ Différents noms ont été donnés à un même paramètre lorsque des conditions ont été définies sur ce paramètre, voir les figures des scénarii pour le détail.

Analyse Laoshan - Guangdong

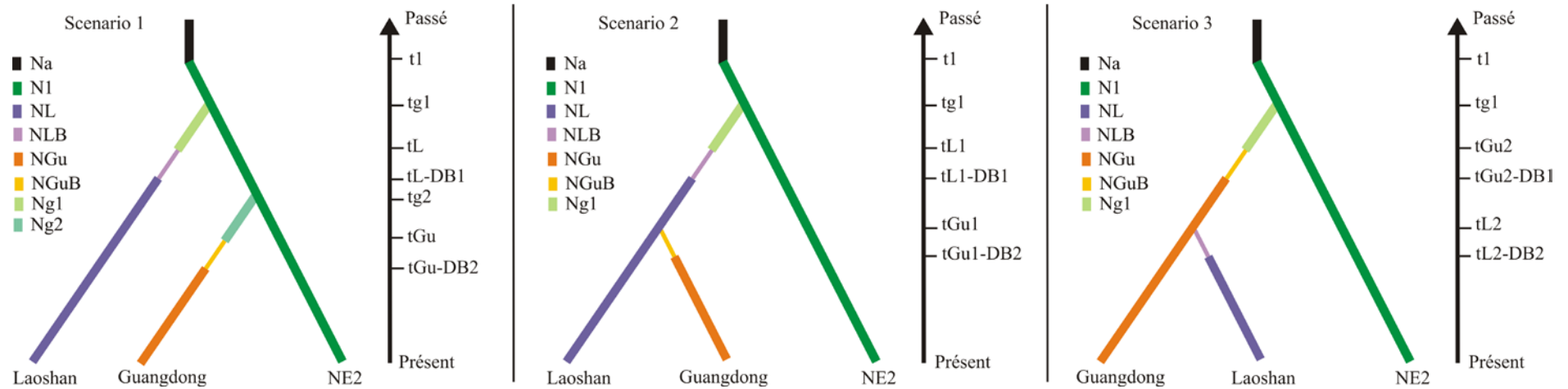


Figure 4 : Représentation des différents scénarii en compétition dans l'analyse ABC réalisée au sein de la Chine dans le but de déterminer si Guangdong était indépendant de Laoshan. Les couleurs dans les scénarii symbolisent différentes tailles efficaces. Voir le Tableau 2 pour la signification des paramètres (tailles efficaces et temps de divergence).

Analyse Kosa - Kasumig3

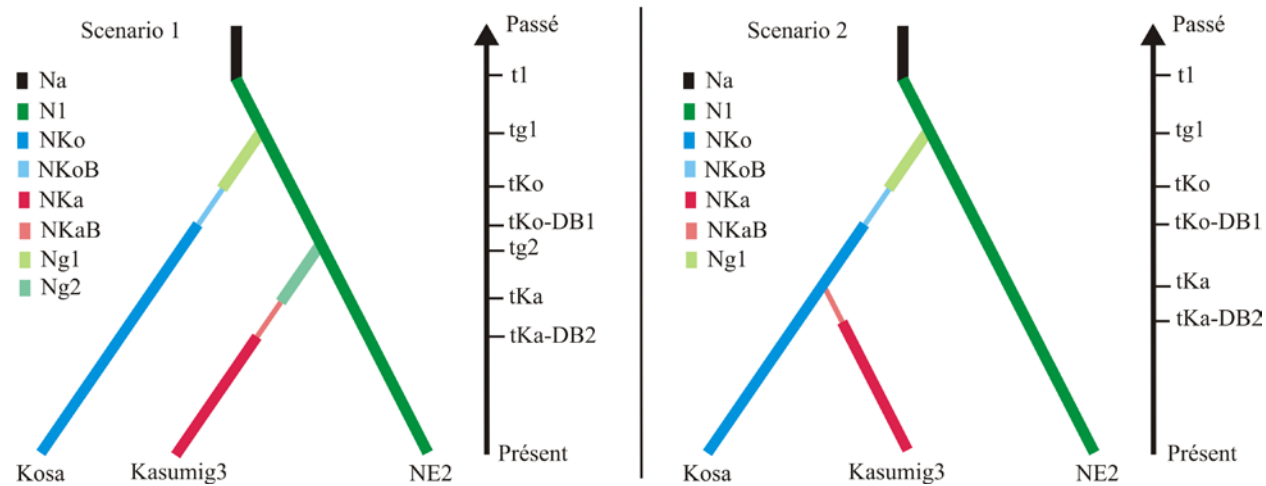
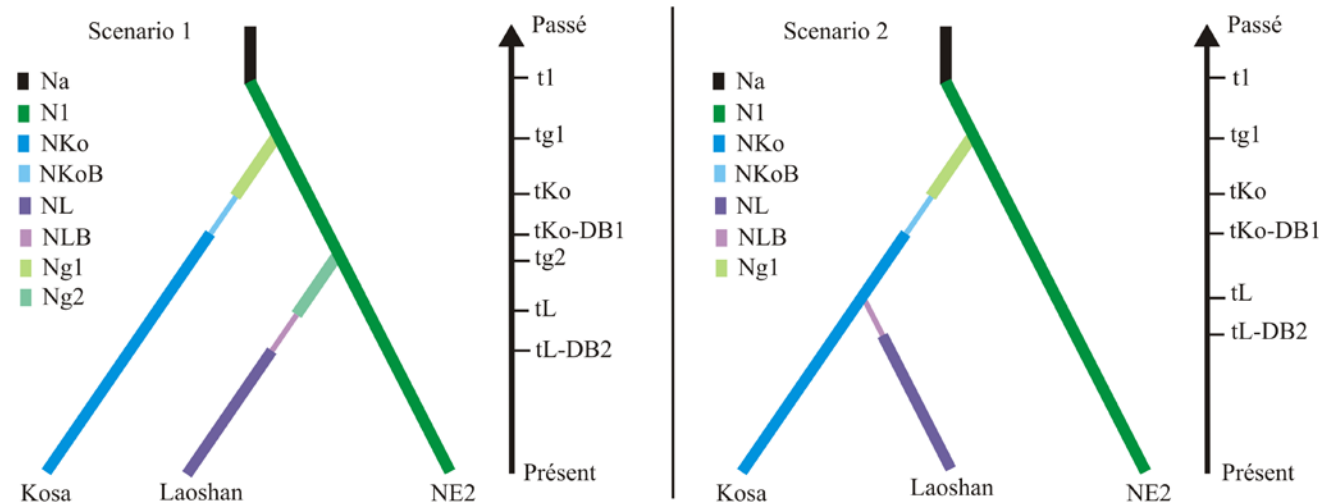


Figure 5 : Représentation des différents scénarii en compétition dans l'analyse ABC réalisée au sein du Japon dans le but de déterminer l'origine de Kasumig3. Les couleurs dans les scénarii symbolisent différentes tailles efficaces. Voir le Tableau 2 pour la signification des paramètres (tailles efficaces et temps de divergence).

Analyse Kosa - Laoshan



Analyse Kosa - Guangdong

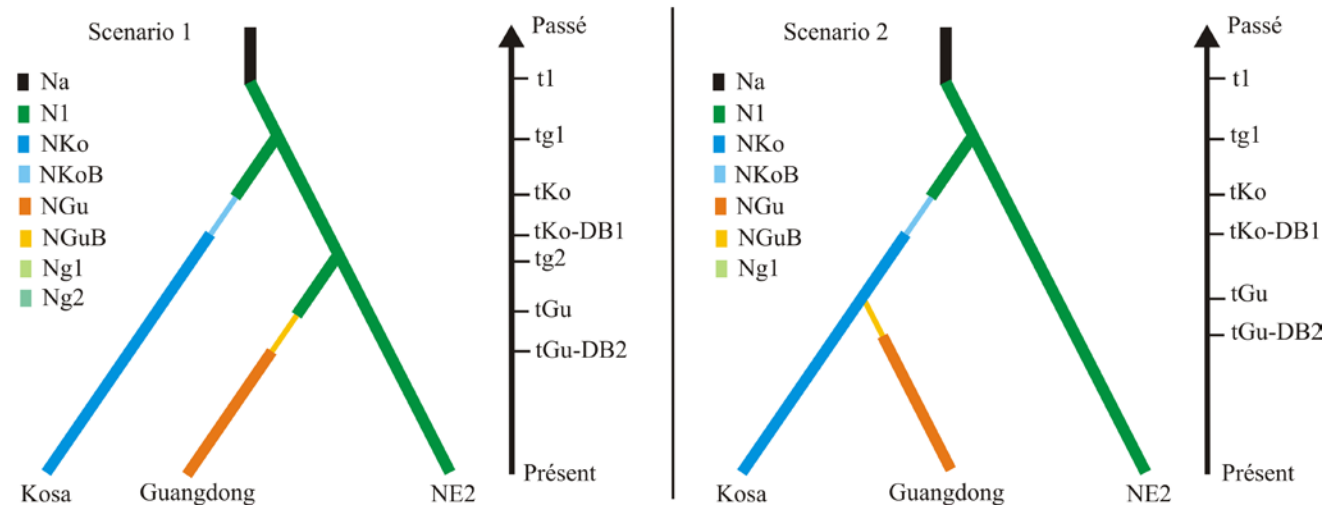


Figure 6 : Représentation des différents scénarii en compétition dans les analyses ABC réalisées entre la Chine et le Japon. Il s'agissait de déterminer si Laoshan et Guangdong venaient du Japon ou des USA. Les couleurs dans les scénarii symbolisent différentes tailles efficaces. Voir le Tableau 2 pour la signification des paramètres (tailles efficaces et temps de divergence).

Analyse Kasumig3 - Guangdong

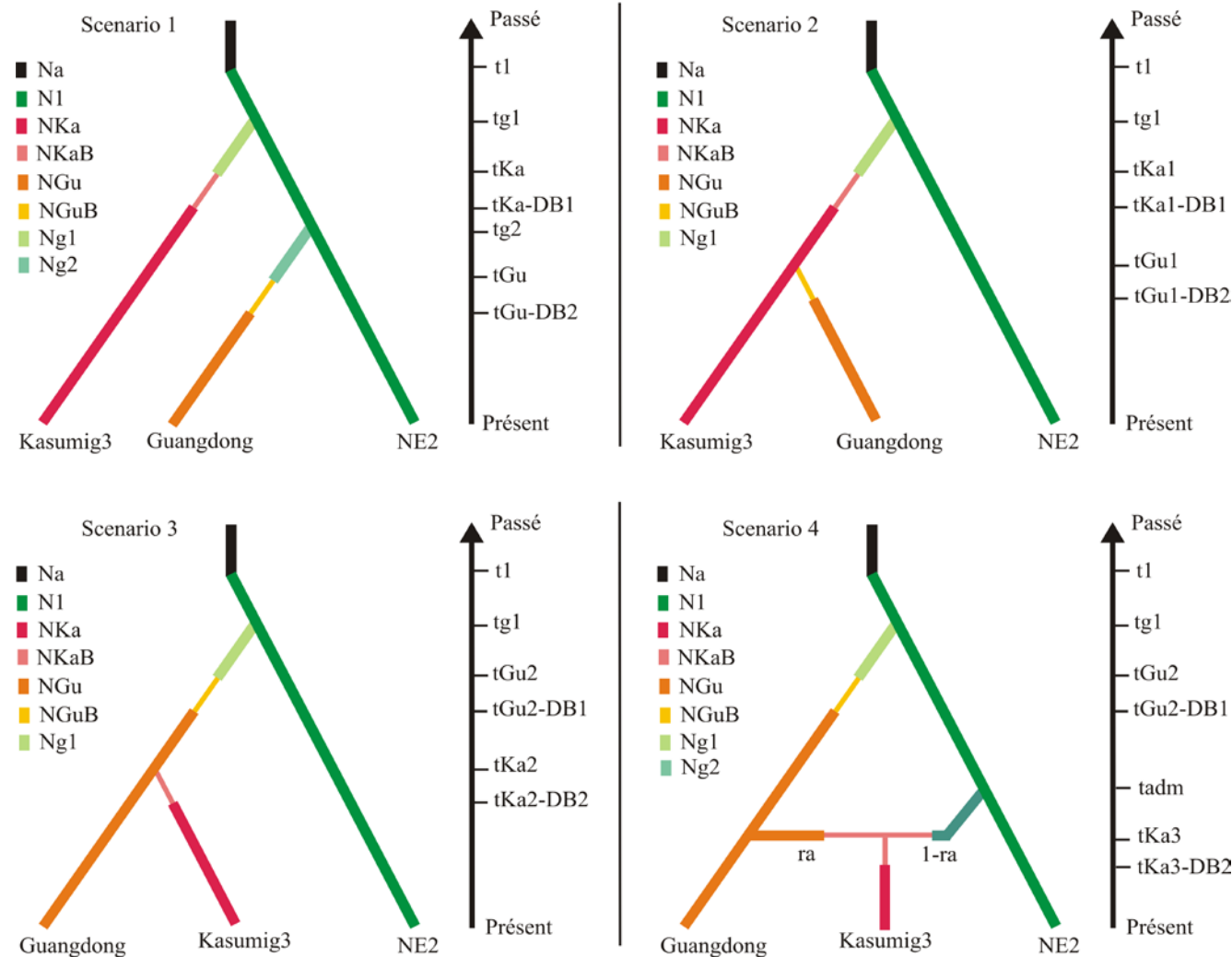


Figure 7 : Représentation des différents scenarii en compétition dans l'analyse ABC réalisée entre Guangdong (Chine) et Kasumig3 (Japon). Ces analyses ont été réalisées au vu des résultats obtenus grâce à l'arbre de distances génétiques et l'analyse de groupement des individus. Dans le scénario 4, où un évènement d'admixture (mélange entre populations) a lieu, le taux d'admixture ra correspond à la contribution génétique de la population envahissante Guangdong. La contribution de la population fantôme native, issue de NE2, correspond à $1-ra$. Les couleurs dans les scenarii symbolisent différentes tailles efficaces. Voir le Tableau 2 pour la signification des paramètres (tailles efficaces et temps de divergence).

Analyse Asie - Europe

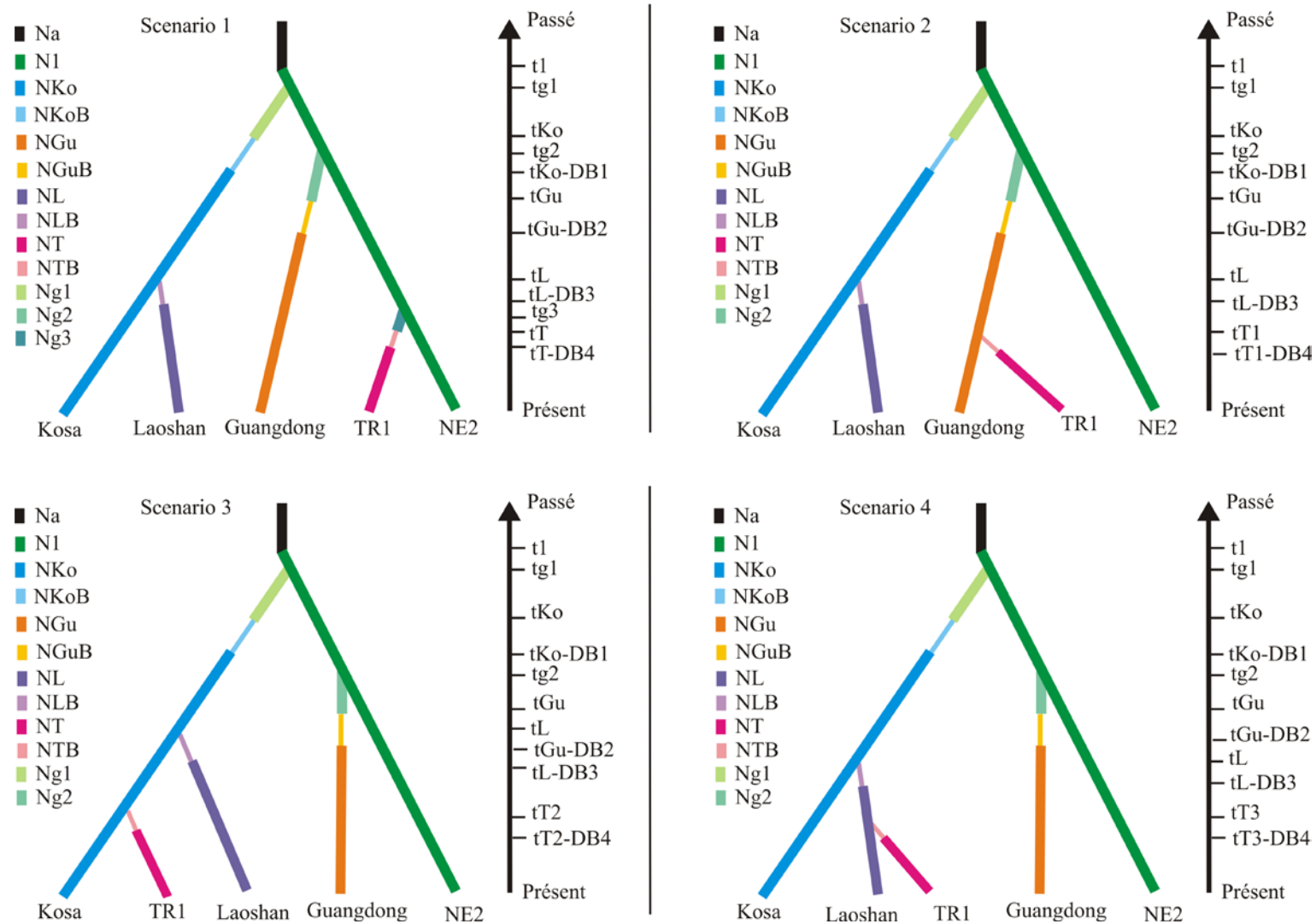


Figure 8 : Représentation des différents scenarii en compétition dans l'analyse ABC réalisée avec toutes les zones envahies, dans le but de déterminer l'origine du Portugal. Les couleurs dans les scenarii symbolisent différentes tailles efficaces. Voir le Tableau 2 pour la signification des paramètres (tailles efficaces et temps de divergence).

Pour chaque analyse ABC effectuée, un million de jeux de données ont été simulés par scénario en compétition. Les statistiques résumées suivantes ont été utilisées pour la comparaison et la sélection de scénarii : le nombre d'allèles moyen, l'hétérozygotie attendue moyenne et la variance des tailles alléliques moyenne par population et par paire de populations, le ratio du nombre d'allèles sur la gamme des tailles alléliques (Garza & Williamson 2001), la vraisemblance moyenne d'assignation des individus de la population i envers la population j (Pascual et al. 2007) et les distances d'allèles partagés entre paires de populations (Chakraborty & Jin 1993). L'adéquation entre les données observées et les *priors* a systématiquement été vérifiée (option "evaluate scenarios and priors" du logiciel). Des probabilités égales ont été définies *a priori* pour chaque scénario. Cela nous a permis d'évaluer l'information contenue dans les données génétiques et leur puissance pour distinguer les différents scénarii, en faisant en partie abstraction des informations historiques (qui sont peut-être erronées). Les probabilités *a posteriori* ont été obtenues par régression logistique en ne conservant que 1% des données simulées les plus proches des données observées (Cornuet et al. 2010; Cornuet et al. 2008). L'option LDA (pour "linear discriminant analysis") qui permet de réaliser une analyse discriminante sur les statistiques résumées avant de procéder à la régression logistique a été utilisée dans les analyses impliquant toutes les zones envahies (Cornuet et al. 2014; Estoup et al. 2012). Les résultats de sélection de modèle pouvant être sensibles aux *priors* (Bertorelle et al. 2010), nous avons réalisé des analyses en modifiant les *priors* précédemment utilisés pour évaluer la robustesse de nos résultats. Ainsi, un deuxième groupe de *priors* a été défini en modifiant la forme des distributions des *priors* par rapport au groupe n°1 (groupe n°2, voir Tableau 2 pour le détail des *priors*) et différents temps de génération ont été testés (voir Tableau 2 pour le détail des *priors*). En plus du changement de *priors*, nous avons évalué la confiance que l'on pouvait avoir dans les résultats obtenus *via* les différentes méthodes permises par l'ABC. Nous avons ainsi (i) vérifié le non-

recoupement des intervalles de confiance associés aux probabilités *a posteriori*, (ii) calculé les erreurs de type I (i.e., la probabilité de ne pas sélectionner un scénario lorsqu'il est "vrai") et de type II (i.e., la probabilité de sélectionner le scénario alors qu'il est "faux", option "Confidence in scenario choice" du logiciel) en simulant 100 nouveaux jeux de données pour chaque scénario considéré et (iii) vérifié l'adéquation entre les données observées et le scénario sélectionné, c'est-à-dire sa capacité à simuler des jeux de données similaires aux données observées (option "model checking" du logiciel). Pour cela, deux nouvelles statistiques résumées ont été ajoutées à celles déjà utilisées pour la comparaison de scénarii, le F_{st} (Weir & Cockerham 1984) et la distance $(d_{\mu})^2$ (Goldstein et al. 1995) entre paires de populations. Lors de cette vérification, la position de chaque statistique observée dans la distribution des statistiques obtenues est estimée *via* les jeux de données simulées, fournissant une probabilité de rejet pour chaque statistique. Un grand nombre de probabilités étant ainsi calculées, nous avons donc ajusté le seuil de significativité par la méthode FDR (Benjamini & Hochberg 1995).

❖ Résultats

Analyses génétiques classiques

Les caractéristiques génétiques des échantillons sont présentées dans le tableau 1. Comme déjà constaté dans les études précédentes, les échantillons américains présentent une diversité génétique faible à modérée dont le niveau est variable entre les échantillons, avec un N_a moyen de 1.27 à 3.87 et une H_e de 0.08 à 0.44. Plus de la moitié des échantillons ne sont pas apparus à HWE et des DL significatifs ont été observés pour 13 tests sur les 1084 réalisés, après correction. Cela est possiblement dû à la présence d'allèles nuls comme déjà mis en évidence dans Mallez et al. (2014). Comme déjà présenté également, le Japon et

Portugal/Madère ont présenté une diversité faible voire extrêmement faible, due à la présence de nombreux marqueurs monomorphes : 9 pour le Japon et 14 pour Portugal/Madère, sur les 15 étudiés. Au Japon, trois échantillons sont apparus complètement "identiques", avec des individus présentant le même allèle fixé à tous les marqueurs (Jap212, Jap308 et Jap120). Au Portugal/Madère, 162 individus sur les 169 génotypés présentent le même génotype multi-locus et seul l'échantillon 128S est différent des autres. Le niveau de diversité génétique rencontrée dans les échantillons chinois s'est avéré comparable au niveau de diversité des deux autres zones envahies. Trois échantillons sur les 4 analysés présentent le même génotype multi-locus pour tous leurs individus. L'échantillon différent, Guangdong, a présenté un allèle fixé différent des trois autres échantillons pour 5 marqueurs, conduisant à 5 marqueurs dialléliques sur les 15 analysés à l'échelle de la Chine, bien que la totalité des marqueurs soient monomorphes en intra-échantillon.

Relation entre les différents échantillons

Certaines localités échantillonnées se sont avérées complètement "identiques" (présentant le même génotype multi-locus pour tous les individus, voir ci-dessus) en Chine, au Japon et au Portugal/Madère. Pour alléger les analyses et simplifier les figures, seule une localité échantillonnée a été conservée parmi les localités "identiques" pour l'analyse Structure, le calcul des vraisemblances d'assignation individuelle et l'arbre de distances génétiques. Les échantillons conservés (parmi les "identiques") sont : Mad23PC pour Madère, TR1 pour le Portugal, Jap212 pour le Japon et Laoshan pour la Chine.

Concernant les tests de différenciation génétique, quasiment tous les échantillons américains sont apparus significativement différenciés entre eux après correction (méthode de Fisher de combinaison des probabilités obtenues sur les différents loci, $p < 10^{-2}$). Seul un test, entre NC2 et GA3, n'est pas apparu significatif (méthode de Fisher, $p = 0.11$). Les valeurs de

Fst corrigées pour les allèles nuls atteignent des valeurs relativement élevées et s'étalent de 0.036 à 0.76 (voir Tableau 3 pour les détails). Les valeurs de D_{est} sont moins élevées et s'échelonnent de 0.0041 à 0.329 (voir Tableau 3). Au Portugal/Madère, tous les échantillons "identiques" se sont avérés significativement différenciés de l'échantillon 128S (méthode de Fisher, $p < 10^{-5}$) avec un Fst moyen corrigé pour les allèles nuls de 0.341 et un D_{est} de 0.0005 (voir Tableau 3). Au Japon, tous les échantillons sont aussi apparus significativement différenciés entre eux, à l'exception des trois "identiques" (méthode de Fisher, $p < 10^{-5}$) avec des valeurs de Fst corrigées pour les allèles nuls très élevées, de 0.627 à 0.995, et des valeurs de D_{est} beaucoup moins élevées, de 0.0042 à 0.0938 (voir Tableau 3). Enfin, en Chine, l'échantillon Guangdong s'est avéré significativement différencié des trois autres échantillons (méthode de Fisher, $p < 10^{-5}$) avec un Fst moyen corrigé pour les allèles nuls très élevé de 0.997 et un D_{est} bien plus faible de 0.1061 (voir Tableau 3). Ce résultat a été confirmé par l'analyse Structure, qui a inféré un groupement pertinent d'individus pour $K=2$, avec Guangdong se séparant des autres échantillons et formant un groupe à part entière (voir Figure 9).

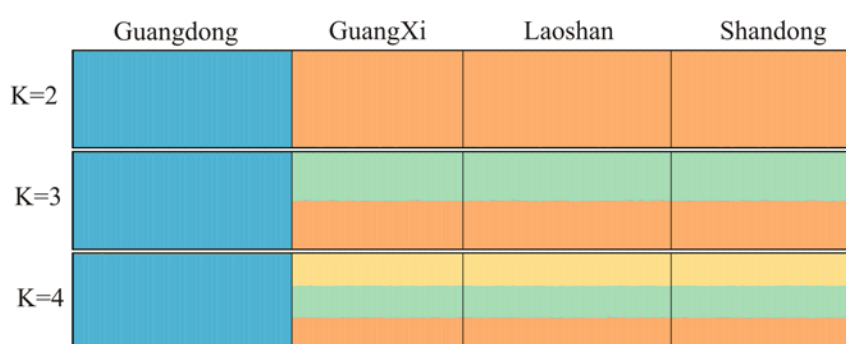


Figure 9 : Représentations graphiques des groupements d'individus obtenus avec Structure, pour les différentes valeurs de K testées, en Chine. Chaque barre verticale correspond à un individu et chaque couleur correspond à un groupe.

Tous les échantillons envahissants sont apparus significativement différenciés entre eux (méthode de Fisher, $p < 10^{-5}$) et significativement différenciés des échantillons natifs (méthode de Fisher, $p < 10^{-5}$). L'échantillon américain NE2 est apparu le moins différencié des échantillons envahissants sur la base du D_{est} (valeur moyenne de 0.034). Un échantillon

différent, NE9, a été observé moins différencié des échantillons envahissants sur la base du F_{st} corrigé pour les allèles nuls (valeur moyenne de 0.469).

L'arbre de distances génétiques (voir Figure 10) indique que l'échantillon NE2 est le plus proche de tous les échantillons des zones envahies. Les échantillons Portugais sont aussi apparus plus proches d'un échantillon japonais que chinois.

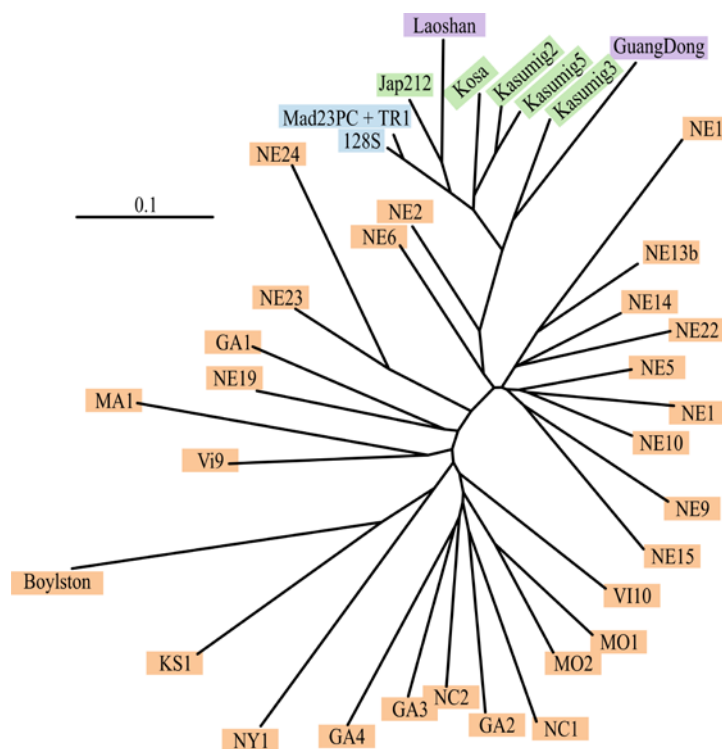


Figure 10 : Arbre de distances génétiques du PWN provenant de son aire native (en orange) et de trois zones envahies (Japon en vert, Chine en mauve, Portugal en bleu) réalisé avec la méthode du Neighbour Joining à partir des distances de Cavalli-Sforza & Edwards (1967), corrigées pour les allèles nuls, obtenues avec FreeNA (Chapuis & Estoup 2007). Parmi les échantillons complètement identiques au Japon, en Chine, au Portugal et à Madère, un seul a été conservé pour simplifier la lecture : Jap212 pour le Japon, Laoshan pour la Chine, TR1 pour le Portugal et Mad23PC pour Madère.

L'analyse Bayésienne de groupement des individus prenant en compte toutes les zones envahies a donné des résultats similaires à ceux obtenus grâce à l'arbre de distances génétiques. L'observation des premières valeurs de K (voir Figure 11, ci-dessous) indique que les échantillons des zones envahies se séparent des échantillons américains dès le premier niveau de structuration, à $K=2$. Les échantillons des zones envahies se divisent ensuite à $K=4$ en deux groupes : un groupe avec Kasumig3 et Guangdong d'un côté et un groupe avec tous les autres échantillons des zones envahies de l'autre.

Enfin, presque tous les échantillons des zones envahies ont été assigné à NE2 avec la vraisemblance d'assignation individuelle moyenne la plus élevée (voir Figure 12, ci-dessous). Seul Guangdong a été assigné à NE5 (voir Figure 12).

Tableau 3 : Mesures de différenciation entre paires d'échantillons des zones native et envahies. Sous la diagonale, le D_{est} de Jost (2008) calculé avec SMOGD (Crawford 2010) et au-dessus de la diagonale, le Fst de Weir & Cockerham (1984), corrigés pour les allèles nuls, obtenu avec FreeNA (Chapuis & Estoup 2007). Code couleur : plus les valeurs sont vertes, plus elles sont faibles et plus elles sont rouges, plus elles sont élevées.

	MO1	MO2	NE1	NE2	NE5	NE6	NE9	NE10	NE12	NE13b	NE14	NE15	NE19	NE22	NE23	NE24	Vi9	Vi10	MA1	NC1	NC2	NY1	GA1	GA2	GA3
MO1		0.065	0.294	0.246	0.279	0.275	0.226	0.328	0.388	0.176	0.221	0.314	0.193	0.250	0.415	0.500	0.334	0.133	0.429	0.127	0.112	0.260	0.208	0.137	0.104
MO2	0.022		0.266	0.194	0.240	0.257	0.194	0.291	0.402	0.173	0.196	0.265	0.151	0.214	0.390	0.504	0.307	0.109	0.387	0.074	0.055	0.217	0.145	0.101	0.074
NE1	0.097	0.089		0.188	0.120	0.168	0.140	0.180	0.493	0.223	0.227	0.238	0.319	0.216	0.505	0.623	0.508	0.333	0.596	0.305	0.253	0.367	0.387	0.268	0.278
NE2	0.067	0.054	0.033		0.145	0.161	0.108	0.250	0.500	0.236	0.188	0.247	0.311	0.110	0.448	0.641	0.474	0.301	0.553	0.238	0.221	0.355	0.330	0.260	0.272
NE5	0.100	0.079	0.009	0.014		0.141	0.087	0.086	0.475	0.190	0.154	0.198	0.292	0.176	0.411	0.567	0.483	0.288	0.582	0.309	0.243	0.342	0.323	0.257	0.278
NE6	0.067	0.087	0.020	0.033	0.020		0.137	0.278	0.502	0.241	0.153	0.369	0.381	0.198	0.503	0.632	0.506	0.328	0.567	0.317	0.281	0.433	0.397	0.286	0.329
NE9	0.071	0.065	0.028	0.023	0.015	0.018		0.153	0.411	0.116	0.108	0.167	0.290	0.121	0.430	0.501	0.398	0.251	0.482	0.206	0.215	0.317	0.296	0.233	0.241
NE10	0.116	0.095	0.016	0.038	0.004	0.031	0.023		0.521	0.187	0.213	0.152	0.341	0.279	0.422	0.547	0.462	0.321	0.609	0.363	0.277	0.376	0.407	0.300	0.336
NE12	0.108	0.141	0.118	0.136	0.106	0.079	0.091	0.119		0.289	0.424	0.599	0.439	0.401	0.662	0.696	0.688	0.457	0.611	0.475	0.427	0.525	0.521	0.463	0.456
NE13b	0.036	0.061	0.055	0.052	0.045	0.035	0.030	0.037	0.049		0.118	0.229	0.244	0.188	0.472	0.556	0.450	0.221	0.504	0.217	0.195	0.303	0.315	0.243	0.214
NE14	0.068	0.064	0.038	0.048	0.012	0.027	0.024	0.025	0.102	0.021		0.275	0.322	0.135	0.417	0.543	0.347	0.216	0.487	0.213	0.185	0.323	0.280	0.256	0.238
NE15	0.112	0.087	0.021	0.048	0.027	0.068	0.040	0.021	0.159	0.053	0.045		0.357	0.294	0.574	0.613	0.456	0.269	0.602	0.284	0.227	0.323	0.383	0.271	0.286
NE19	0.068	0.036	0.064	0.080	0.058	0.096	0.089	0.074	0.107	0.093	0.094	0.059		0.306	0.481	0.573	0.491	0.283	0.541	0.270	0.167	0.310	0.225	0.219	0.174
NE22	0.064	0.072	0.049	0.019	0.020	0.044	0.034	0.059	0.119	0.034	0.028	0.071	0.073		0.476	0.611	0.441	0.285	0.510	0.217	0.205	0.311	0.297	0.268	0.247
NE23	0.120	0.106	0.048	0.052	0.033	0.062	0.080	0.044	0.132	0.095	0.060	0.088	0.069	0.091		0.594	0.576	0.474	0.715	0.472	0.391	0.580	0.579	0.391	0.504
NE24	0.171	0.163	0.107	0.145	0.080	0.128	0.155	0.091	0.218	0.145	0.118	0.133	0.103	0.177	0.023		0.711	0.570	0.760	0.524	0.474	0.589	0.668	0.435	0.511
Vi9	0.109	0.095	0.099	0.105	0.087	0.119	0.123	0.069	0.286	0.122	0.048	0.085	0.122	0.113	0.096	0.151		0.317	0.604	0.325	0.258	0.418	0.450	0.338	0.352
Vi10	0.054	0.055	0.117	0.103	0.096	0.097	0.102	0.095	0.170	0.071	0.064	0.079	0.111	0.102	0.110	0.178	0.072		0.378	0.118	0.094	0.236	0.229	0.158	0.113
MA1	0.183	0.146	0.236	0.217	0.211	0.176	0.230	0.215	0.199	0.198	0.172	0.089	0.167	0.212	0.222	0.304	0.155	0.099		0.405	0.379	0.474	0.541	0.418	0.442
NC1	0.064	0.035	0.118	0.099	0.133	0.120	0.103	0.135	0.216	0.068	0.069	0.092	0.112	0.072	0.152	0.200	0.080	0.042	0.149		0.057	0.198	0.224	0.106	0.063
NC2	0.061	0.020	0.073	0.080	0.084	0.111	0.096	0.085	0.201	0.065	0.062	0.069	0.050	0.064	0.101	0.146	0.055	0.047	0.143	0.012		0.157	0.129	0.091	0.016
NY1	0.154	0.108	0.134	0.136	0.152	0.217	0.192	0.150	0.254	0.164	0.142	0.112	0.167	0.143	0.189	0.245	0.139	0.109	0.207	0.089	0.080		0.248	0.245	0.125
GA1	0.090	0.048	0.097	0.090	0.078	0.112	0.118	0.122	0.115	0.128	0.088	0.114	0.051	0.099	0.122	0.177	0.093	0.092	0.168	0.110	0.051	0.094		0.234	0.130
GA2	0.065	0.056	0.090	0.103	0.091	0.084	0.111	0.093	0.202	0.097	0.087	0.091	0.076	0.104	0.095	0.112	0.090	0.061	0.148	0.027	0.039	0.124	0.107		0.053
GA3	0.046	0.019	0.074	0.077	0.081	0.105	0.104	0.116	0.187	0.072	0.081	0.085	0.048	0.081	0.097	0.110	0.086	0.038	0.175	0.018	0.002	0.057	0.053	0.010	
GA4	0.054	0.049	0.088	0.094	0.094	0.115	0.088	0.093	0.192	0.057	0.073	0.100	0.091	0.071	0.126	0.162	0.083	0.062	0.212	0.076	0.032	0.112	0.094	0.057	0.009
KS1	0.066	0.047	0.145	0.108	0.145	0.160	0.125	0.160	0.204	0.108	0.106	0.150	0.139	0.134	0.193	0.226	0.160	0.096	0.218	0.066	0.050	0.119	0.096	0.095	0.038
Boylston	0.186	0.197	0.300	0.303	0.329	0.315	0.320	0.288	0.265	0.254	0.300	0.308	0.211	0.314	0.233	0.292	0.262	0.220	0.222	0.252	0.210	0.253	0.250	0.227	0.194
Mad23PC	0.135	0.107	0.059	0.022	0.042	0.053	0.052	0.082	0.216	0.140	0.085	0.084	0.111	0.053	0.107	0.213	0.127	0.170	0.270	0.166	0.139	0.184	0.118	0.152	0.153
Mad24C	0.134	0.105	0.058	0.022	0.041	0.052	0.051	0.081	0.216	0.139	0.084	0.084	0.110	0.052	0.107	0.212	0.126	0.169	0.270	0.164	0.138	0.182	0.118	0.151	0.151
E182	0.135	0.107	0.059	0.022	0.042	0.053	0.052	0.082	0.216	0.140	0.085	0.085	0.111	0.054	0.107	0.213	0.127	0.170	0.270	0.166	0.140	0.184	0.118	0.153	0.153
128S	0.135	0.104	0.059	0.022	0.042	0.053	0.052	0.082	0.217	0.140	0.085	0.085	0.112	0.054	0.107	0.213	0.116	0.168	0.270	0.159	0.130	0.184	0.116	0.149	0.149
E1069	0.135	0.107	0.059	0.022	0.042	0.053	0.052	0.082	0.216	0.140	0.085	0.085	0.111	0.054	0.107	0.213	0.127	0.170	0.270	0.166	0.140	0.184	0.118	0.153	0.153
AM2	0.136	0.108	0.059	0.023	0.042	0.053	0.053	0.082	0.217	0.141	0.086	0.085	0.111	0.054	0.107	0.213	0.127	0.171	0.270	0.166	0.140	0.185	0.119	0.153	0.153
Comporta	0.137	0.108	0.060	0.023	0.042	0.053	0.053	0.082	0.217	0.141	0.086	0.085	0.112	0.054	0.107	0.213	0.127	0.171	0.270	0.167	0.141	0.186	0.119	0.153	0.154
TR1	0.137	0.108	0.060	0.023	0.042	0.054	0.053	0.082	0.217	0.141	0.086	0.085	0.112	0.055	0.107	0.213	0.127	0.171	0.270	0.167	0.141	0.186	0.119	0.153	0.154
TR2	0.136	0.108	0.060	0.023	0.042	0.053	0.053	0.082	0.217	0.141	0.086	0.085	0.112	0.054	0.107	0.213	0.127	0.171	0.270	0.167	0.140	0.186	0.119	0.153	0.154
Jap308	0.135	0.135	0.044	0.041	0.060	0.079	0.078	0.102	0.217	0.159	0.094	0.096	0.096	0.063	0.107	0.213	0.121	0.173	0.270	0.141	0.110	0.166	0.116	0.130	0.102
Jap212	0.135	0.135	0.044	0.041	0.060	0.079	0.078	0.102	0.217	0.159	0.094	0.096	0.096	0.063	0.107	0.213	0.121	0.173	0.270	0.141	0.110	0.166	0.116	0.130	0.102
Jap120	0.134	0.135	0.044	0.041	0.059	0.079	0.078	0.102	0.217	0.158	0.094	0.096	0.096	0.063	0.107	0.213	0.121	0.173	0.270	0.141	0.110	0.166	0.116	0.130	0.102
Kasumig2	0.125	0.134	0.060	0.023	0.043	0.054	0.053	0.082	0.217	0.141	0.086	0.083	0.124	0.057	0.092	0.213	0.110	0.155	0.206	0.153	0.129	0.184	0.129	0.146	0.121
Kasumig3	0.086	0.079	0.088	0.045	0.059	0.044	0.074	0.091	0.096	0.102	0.109	0.115	0.082	0.065	0.068	0.152	0.185	0.103	0.201	0.121	0.133	0.212	0.129	0.088	0.101
Kasumig5	0.090	0.116	0.092	0.035	0.067	0.071	0.083	0.104	0.160	0.130	0.106	0.122	0.082	0.073	0.058	0.158	0.135	0.150	0.205	0.146	0.112	0.184	0.120	0.132	0.101
Kosa	0.132	0.132	0.058	0.021	0.036	0.049	0.045	0.072	0.217	0.135	0.077	0.073	0.121	0.053	0.088	0.207	0.107	0.133	0.153	0.160	0.138	0.192	0.134	0.165	0.147
GuangDong	0.154	0.168	0.138	0.085	0.106	0.071	0.101	0.164	0.160	0.165	0.126	0.163	0.172	0.095	0.155	0.275	0.206	0.140	0.159	0.165	0.190	0.255	0.190	0.182	0.194
GuangXi	0.153	0.176	0.094	0.050	0.072	0.087	0.068	0.107	0.227	0.147	0.123	0.110	0.149	0.090	0.155	0.279	0.178	0.227	0.340	0.216	0.179	0.257	0.165	0.192	

Suite du **Tableau 3.**

	GA4	KS1	Boylston	Mad23PC	Mad24C	E182	128S	E1069	AM2	Comporta	TR1	TR2	Jap308	Jap212	Jap120	Kasumig2	Kasumig3	Kasumig5	Kosa	GuangDong	GuangXi	Laoshan	ShanDong
MO1	0.086	0.209	0.322	0.489	0.450	0.496	0.483	0.501	0.544	0.576	0.586	0.573	0.561	0.571	0.553	0.589	0.442	0.531	0.533	0.645	0.617	0.643	0.628
MO2	0.092	0.189	0.318	0.453	0.403	0.462	0.444	0.468	0.521	0.560	0.572	0.556	0.569	0.580	0.559	0.616	0.431	0.591	0.532	0.673	0.644	0.673	0.656
NE1	0.263	0.347	0.478	0.606	0.549	0.615	0.581	0.619	0.676	0.713	0.724	0.710	0.669	0.682	0.659	0.745	0.614	0.787	0.608	0.819	0.779	0.802	0.790
NE2	0.257	0.309	0.453	0.414	0.350	0.425	0.354	0.432	0.496	0.542	0.557	0.539	0.611	0.625	0.600	0.548	0.427	0.608	0.416	0.758	0.695	0.724	0.708
NE5	0.256	0.315	0.461	0.572	0.506	0.583	0.543	0.589	0.652	0.694	0.706	0.690	0.722	0.734	0.712	0.735	0.557	0.761	0.574	0.818	0.768	0.793	0.779
NE6	0.300	0.369	0.494	0.561	0.518	0.570	0.524	0.572	0.621	0.654	0.665	0.651	0.695	0.705	0.687	0.685	0.514	0.721	0.556	0.742	0.753	0.775	0.763
NE9	0.202	0.270	0.415	0.381	0.341	0.388	0.362	0.391	0.436	0.469	0.480	0.466	0.514	0.525	0.505	0.497	0.429	0.570	0.392	0.611	0.525	0.553	0.537
NE10	0.291	0.366	0.488	0.609	0.569	0.616	0.595	0.619	0.662	0.693	0.702	0.690	0.704	0.713	0.696	0.720	0.621	0.737	0.618	0.809	0.737	0.759	0.747
NE12	0.436	0.434	0.486	0.768	0.741	0.774	0.762	0.774	0.806	0.826	0.832	0.824	0.818	0.824	0.812	0.844	0.671	0.817	0.773	0.831	0.842	0.857	0.848
NE13b	0.177	0.265	0.394	0.582	0.527	0.591	0.574	0.597	0.652	0.690	0.701	0.687	0.685	0.696	0.675	0.722	0.565	0.715	0.617	0.765	0.723	0.750	0.735
NE14	0.205	0.276	0.440	0.511	0.471	0.518	0.493	0.521	0.566	0.598	0.609	0.595	0.607	0.617	0.599	0.629	0.542	0.653	0.516	0.688	0.672	0.696	0.682
NE15	0.256	0.340	0.504	0.598	0.552	0.606	0.578	0.612	0.657	0.690	0.700	0.687	0.690	0.700	0.682	0.715	0.631	0.759	0.585	0.815	0.745	0.768	0.755
NE19	0.203	0.278	0.386	0.621	0.567	0.630	0.619	0.635	0.688	0.724	0.735	0.721	0.687	0.698	0.677	0.768	0.591	0.715	0.659	0.818	0.770	0.793	0.780
NE22	0.230	0.260	0.420	0.435	0.378	0.446	0.392	0.451	0.512	0.555	0.568	0.551	0.566	0.579	0.555	0.609	0.446	0.639	0.482	0.707	0.671	0.701	0.684
NE23	0.433	0.493	0.535	0.816	0.796	0.819	0.785	0.821	0.844	0.860	0.865	0.859	0.853	0.859	0.850	0.863	0.715	0.824	0.754	0.906	0.894	0.904	0.898
NE24	0.486	0.507	0.562	0.899	0.879	0.903	0.880	0.904	0.922	0.933	0.936	0.932	0.928	0.932	0.926	0.942	0.833	0.934	0.858	0.951	0.946	0.953	0.949
V19	0.308	0.410	0.519	0.740	0.705	0.745	0.718	0.749	0.784	0.808	0.815	0.805	0.790	0.797	0.783	0.814	0.754	0.830	0.708	0.873	0.848	0.863	0.855
VII0	0.120	0.225	0.356	0.595	0.542	0.605	0.590	0.610	0.664	0.701	0.712	0.697	0.687	0.698	0.678	0.717	0.568	0.709	0.600	0.740	0.754	0.778	0.764
MA1	0.441	0.420	0.462	0.803	0.771	0.809	0.796	0.811	0.843	0.863	0.869	0.861	0.854	0.860	0.849	0.860	0.759	0.858	0.753	0.860	0.885	0.898	0.891
NC1	0.113	0.180	0.338	0.547	0.480	0.559	0.540	0.565	0.632	0.676	0.688	0.671	0.638	0.651	0.626	0.698	0.531	0.685	0.595	0.740	0.739	0.767	0.751
NC2	0.064	0.158	0.305	0.495	0.445	0.503	0.487	0.509	0.561	0.600	0.611	0.596	0.550	0.563	0.541	0.624	0.508	0.599	0.549	0.685	0.665	0.692	0.676
NY1	0.165	0.186	0.350	0.630	0.555	0.642	0.631	0.648	0.714	0.755	0.767	0.752	0.716	0.729	0.705	0.781	0.664	0.780	0.678	0.831	0.802	0.826	0.812
GA1	0.185	0.182	0.389	0.702	0.625	0.714	0.683	0.721	0.781	0.817	0.826	0.814	0.798	0.810	0.790	0.852	0.685	0.843	0.705	0.896	0.865	0.883	0.873
GA2	0.094	0.215	0.340	0.522	0.472	0.531	0.520	0.536	0.589	0.628	0.639	0.624	0.581	0.593	0.571	0.662	0.492	0.647	0.589	0.715	0.682	0.709	0.693
GA3	0.035	0.127	0.295	0.652	0.567	0.666	0.643	0.674	0.742	0.783	0.794	0.780	0.725	0.739	0.713	0.798	0.614	0.777	0.676	0.849	0.824	0.847	0.834
GA4		0.154	0.285	0.525	0.464	0.536	0.521	0.542	0.604	0.647	0.659	0.642	0.600	0.613	0.588	0.669	0.549	0.659	0.577	0.744	0.663	0.694	0.676
KS1	0.058		0.246	0.525	0.466	0.535	0.529	0.542	0.600	0.642	0.654	0.638	0.619	0.633	0.609	0.672	0.558	0.658	0.590	0.737	0.691	0.720	0.703
Boylston	0.189	0.157		0.625	0.592	0.631	0.631	0.635	0.671	0.698	0.706	0.696	0.682	0.690	0.675	0.701	0.610	0.663	0.646	0.741	0.716	0.737	0.725
Mad23PC	0.160	0.216	0.366		nan	nan	0.295	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.744	0.995	0.726	0.996	0.992	0.992	0.992
Mad24C	0.157	0.214	0.366	nan	nan	nan	0.243	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.718	0.995	0.701	0.996	0.992	0.992	0.992
E182	0.160	0.217	0.367	nan	nan		0.304	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.750	0.995	0.730	0.996	0.992	0.992	0.992
128S	0.154	0.216	0.367	0.001	0.001	0.001		0.312	0.362	0.398	0.414	0.398	0.915	0.919	0.912	0.913	0.687	0.942	0.667	0.962	0.926	0.934	0.930
E1069	0.160	0.217	0.367	nan	nan	nan	0.001	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.752	0.995	0.733	0.996	0.992	0.992	0.992
AM2	0.161	0.218	0.367	nan	nan	nan	0.001	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.782	0.995	0.760	0.996	0.992	0.992	0.992
Comporta	0.161	0.218	0.367	nan	nan	nan	0.001	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.801	0.995	0.779	0.996	0.992	0.992	0.992
TR1	0.161	0.218	0.367	nan	nan	nan	0.001	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.809	0.995	0.786	0.996	0.992	0.992	0.992
TR2	0.161	0.218	0.367	nan	nan	nan	0.001	nan	nan	nan	nan	nan	0.992	0.992	0.992	0.992	0.800	0.995	0.778	0.996	0.992	0.992	0.992
Jap308	0.132	0.218	0.357	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017		nan	nan	0.992	0.883	0.995	0.821	0.997	0.992	0.993	0.993
Jap212	0.132	0.218	0.357	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017	nan		nan	0.993	0.886	0.995	0.827	0.997	0.992	0.993	0.993
Jap120	0.131	0.218	0.357	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017	nan	nan		0.992	0.879	0.995	0.817	0.997	0.992	0.992	0.992
Kasumig2	0.141	0.216	0.302	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017		0.864	0.985	0.627	0.996	0.992	0.992	0.992
Kasumig3	0.148	0.205	0.271	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.054		0.835	0.746	0.866	0.895	0.905	0.899
Kasumig5	0.130	0.197	0.236	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.038	0.004	0.033		0.789	0.996	0.995	0.995	0.995
Kosa	0.147	0.215	0.306	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.036	0.036	0.036	0.004	0.058	0.017		0.875	0.828	0.841	0.834
GuangDong	0.222	0.297	0.366	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.106	0.106	0.106	0.068	0.040	0.068	0.041		0.997	0.997	0.997
GuangXi	0.154	0.265	0.373	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.094	0.038	0.034	0.106		nan	nan
Laoshan	0.154	0.265	0.373	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.094	0.038	0.034	0.106	nan		nan
ShanDong	0.154	0.265	0.373	0.017	0.017	0.017	0.020	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.094	0.038	0.034	0.106	nan	nan	

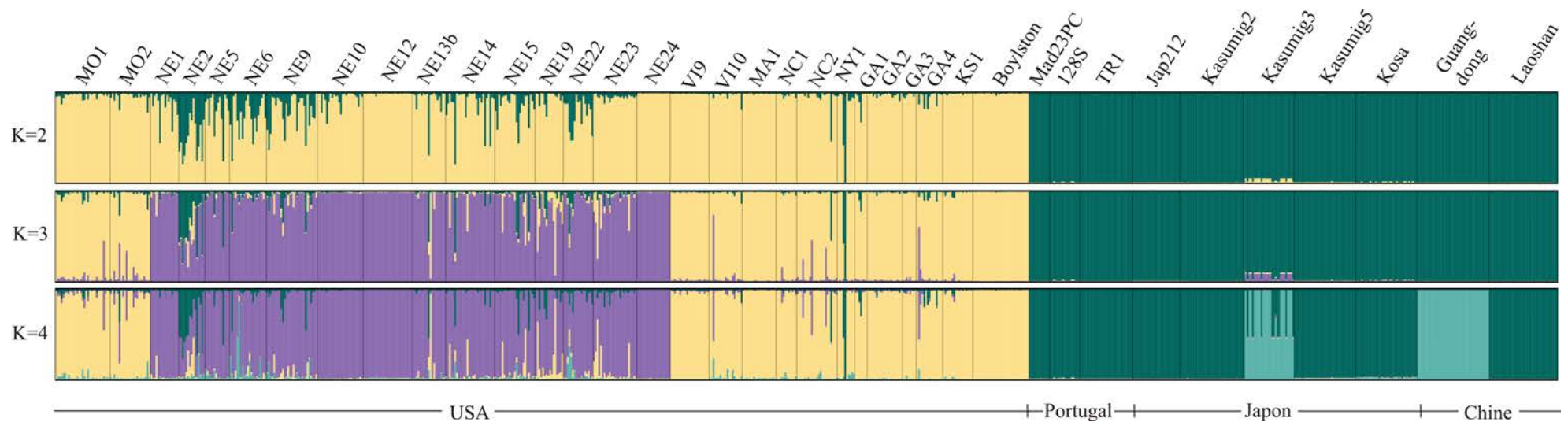


Figure 11 : Représentations graphiques des groupements d'individus obtenus avec Structure, pour les premières valeurs de K testées, avec toutes les zones, natives et envahies, étudiées ici. Chaque barre verticale correspond à un individu et chaque couleur correspond à un groupe. Parmi les échantillons complètement identiques au Japon, en Chine, au Portugal et à Madère, un seul a été conservé pour simplifier la lecture : Jap212 pour le Japon, Laoshan pour la Chine, TR1 pour le Portugal et Mad23PC pour Madère.

Échantillons américains

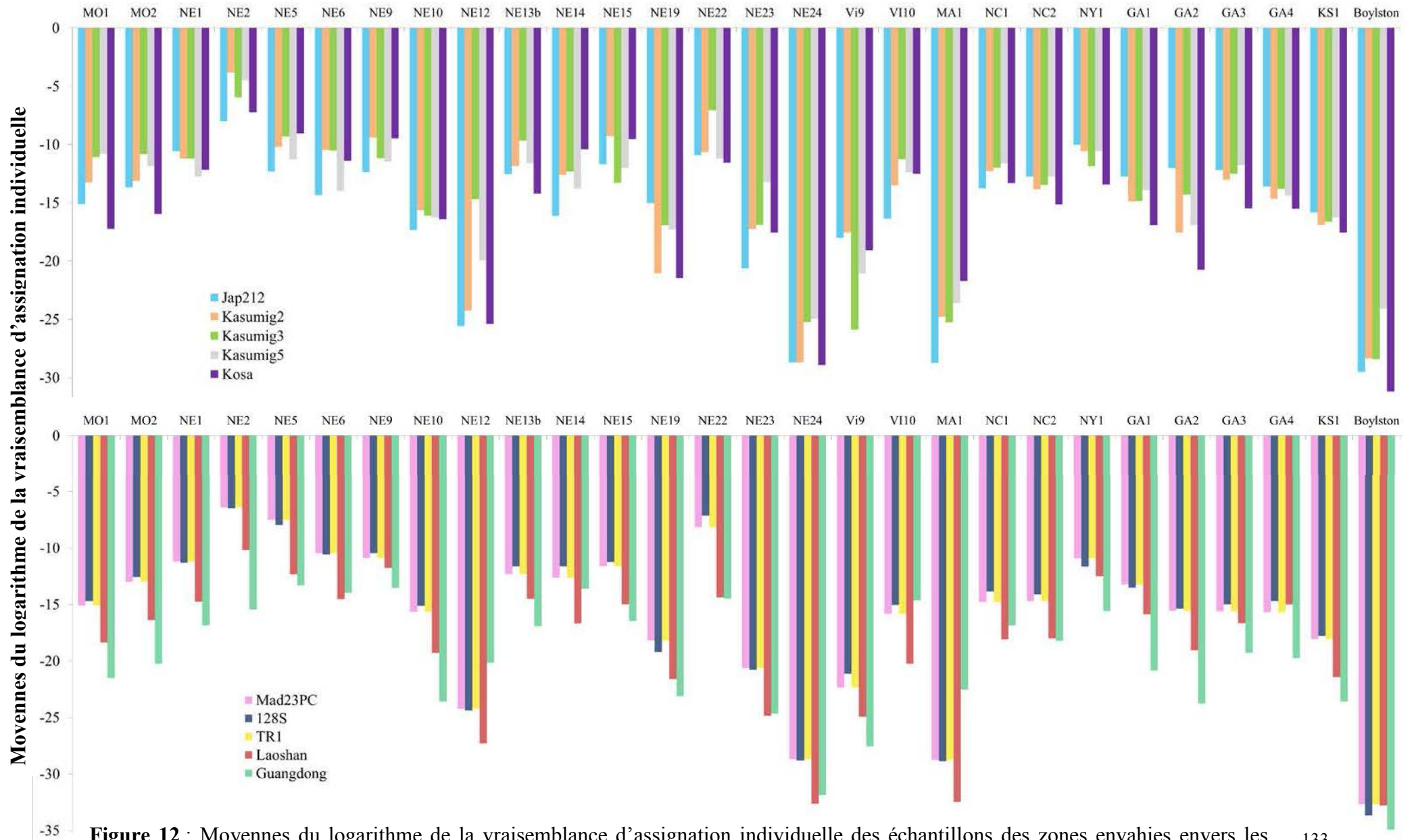


Figure 12 : Moyennes du logarithme de la vraisemblance d'assignation individuelle des échantillons des zones envahies envers les échantillons américains. Parmi les échantillons complètement identiques au Japon, en Chine, au Portugal et à Madère, un seul a été conservé pour simplifier la lecture : Jap212 pour le Japon, Laoshan pour la Chine, TR1 pour le Portugal et Mad23PC pour Madère.

Analyses ABC

Les analyses ABC préliminaires nous ont permis de déterminer que la meilleure façon de modéliser les échantillons américains natifs (parmi les modèles considérés) ne considérait pas de goulot d'étranglement entre la population ancestrale et chacun d'eux. Également, elles ont conduit à baisser l'intervalle de la distribution *a priori* du taux de mutation moyen à $[10^{-6} ; 10^{-4}]$, intervalle qui a été utilisé dans les analyses de comparaison de scénarii.

L'information contenue dans les données génétiques nous a permis de sélectionner nettement des scénarii dans la grande majorité des analyses ABC réalisées (voir Tableaux 4 et 5). La comparaison de scénarii réalisées en Chine a mis en évidence que la population envahissante de Laoshan était indépendant de celle de Guangdong, suggérant deux introductions indépendantes en Chine (voir Tableau 4 pour résultats des comparaisons). Ce résultat est apparu robuste à un changement de *priors*, un résultat similaire ayant été obtenu avec le groupe de *priors* n°2 (voir Tableau 4). Au Japon, le scénario sélectionné sur la base des probabilités *a posteriori* a suggéré que Kasumig3 était indépendant de Kosa quel que soit le groupe de *priors* considéré, indiquant également deux introductions indépendantes au Japon (voir Tableau 4). Les analyses réalisées dans le but d'éclaircir les relations entre la Chine et le Japon ont révélé que : (i) Kosa était probablement à l'origine de Laoshan, suggérant une introduction à partir du Japon en Chine (résultat robuste aux modifications de *priors*, voir Tableau 4), et que (ii) Guangdong était indépendant de Kosa, indiquant que la deuxième introduction suggérée en Chine est possiblement d'origine américaine (résultat robuste aux modifications de *priors*, voir Tableau 4). Les relations entre Guangdong et Kasumig3 n'ont pas été déterminées précisément par les comparaisons de scénarii. Les résultats sont notamment apparus sensibles aux *priors* et des résultats différents ont été obtenus dépendamment du groupe de *priors* utilisé. Avec le groupe de *priors* n°1, le scénario avec évènement d'admixture (mélange entre populations) est apparu le plus probable mais la

probabilité *a posteriori* s'est avérée faible avec de plus une erreur de type I très élevée (voir Tableau 4). Après examen plus précis de l'erreur de type I, nous avons observé que lorsque l'on simulait le scénario avec admixture, tous les scénarii comparés avaient une chance équiprobable d'être sélectionné. Cela suggère que nos données génétiques ne nous permettent pas de différencier les 4 scénarii en compétition dans cette analyse. Avec le groupe de priors n°2, nous n'avons même pas pu sélectionner de scénario. Le scénario 1 et le scénario 4 ont obtenu des probabilités *a posteriori* très proches dont les intervalles de confiance se chevauchaient (voir Tableau 4). Au vu de ce constat, Kasumig3 n'a pas été inclus dans le scénario d'invasion asiatique, ne pouvant pas être placé de manière précise.

Les analyses ABC regroupant toutes les zones envahies étudiées ici ont mis en évidence que les populations portugaises ne venaient pas d'Asie. Une origine américaine est apparue plus probable (voir Tableau 5). Nous avons obtenu ce même résultat avec le groupe de priors n°2. Ce résultat est également apparu robuste aux modifications de priors relatifs au temps de générations par an. En effet, le nombre de 30 générations par an ne correspond pas à une connaissance précise mais à une déduction (comme expliqué dans la partie "Matériel et méthodes"). Nous avons donc testé différents temps de générations pour évaluer l'impact que pouvait avoir cette déduction sur les résultats, si elle sous-estimait (test à 45 générations/an) ou si elle surestimait (test à 15 générations par an) le temps de générations réel (mais inconnu de manière précise) du PWN. Nous avons constaté que le même résultat avait été obtenu quel que soit le nombre de générations par an considéré (voir Tableau 5). Bien que Kasumig3 n'ait pu être inclus dans le scénario d'invasion asiatique, nous avons néanmoins vérifié par une analyse ABC indépendante si cet échantillon pouvait être à l'origine du Portugal. Il s'est avéré que le scénario avec une origine américaine pour le Portugal est apparu plus probable que le scénario modélisant une origine japonaise (résultats non montrés).

Finalement, les vérifications réalisées (calculs d'erreurs et adéquation données-scenario sélectionné) nous ont confortés dans les résultats obtenus. En effet, les erreurs de type I et de type II sont apparues faibles pour toutes les analyses effectuées (l'analyse Kasumig3 – Guangdong mise à part), ce qui démontre que les scenarii en compétition étaient bien différenciables. Également, les scenarii sélectionnés dans les différentes analyses et leurs *posteriors* se sont avérés capables de générer des données proches des données observées. Aucune statistique observée n'est apparue significativement en marge de la gamme obtenue par celles simulées (données non montrées).

Tableau 4 : Résultats des comparaisons de scénarii obtenus par la méthode ABC et réalisées dans le but de clarifier l'histoire de l'invasion asiatique. NE2 = échantillon américain natif, Kosa = échantillon japonais, Laoshan = échantillon chinois et Guangdong = échantillon chinois. Les probabilités *a posteriori* et leur intervalle de confiance ont été obtenus par régression logistique. Le scénario sélectionné sur la base des probabilités *a posteriori* et de leur intervalle de confiance (non-chevauchants) est surligné en vert. Les erreurs de type I et de type II ont été calculé pour ce scénario.

Analyse et scénario	Probabilité <i>a posteriori</i>	Intervalle de confiance	Erreur type I	Erreur type II
Analyses - Chine				
Groupe de priors n°1 - 30 générations/an				
NE2 -> Laoshan ; NE2 -> Guangdong	0.9415	[0.9304 ; 0.9526]	0,13	0,035
NE2 -> Laoshan -> Guangdong	0.0135	[0.0093 ; 0.0176]		
NE2 -> Guangdong -> Laoshan	0.0451	[0.0354 ; 0.0548]		
Groupe de priors n°2 - 30 générations/an				
NE2 -> Laoshan ; NE2 -> Guangdong	0.7252	[0.6806 ; 0.7697]	0,11	0,09
NE2 -> Laoshan -> Guangdong	0.0716	[0.0526 ; 0.0905]		
NE2 -> Guangdong -> Laoshan	0.2032	[0.1646 ; 0.2419]		
Analyses - Japon				
Groupe de priors n°1 - 30 générations/an				
NE2 -> Kosa ; NE2 -> Kasumig3	0.9746	[0.9662 ; 0.9838]	0,17	0,05
NE2 -> Kosa -> Kasumig3	0.0254	[0.0170 ; 0.0338]		
Groupe de priors n°2 - 30 générations/an				
NE2 -> Kosa ; NE2 -> Kasumig3	0.9965	[0.9951 ; 0.9980]	0,07	0,04
NE2 -> Kosa -> Kasumig3	0.0035	[0.0020 ; 0.0049]		
Analyses - Japon / Chine				
Groupe de priors n°1 - 30 générations/an				
NE2 -> Kosa ; NE2 -> Laoshan	0.2537	[0.1940 ; 0.3135]		
NE2 -> Kosa -> Laoshan	0.7463	[0.6865 ; 0.8060]	0,04	0,08
NE2 -> Kosa ; NE2 -> Guangdong	0.8576	[0.8166 ; 0.8986]	0,08	0,09
NE2 -> Kosa -> Guangdong	0.1424	[0.1014 ; 0.1834]		
NE2 -> Kasumig3 ; NE2 -> Guangdong	0,3405	[0.2963 ; 0.3847]		
NE2 -> Kasumig3 -> Guangdong	0,1676	[0.1347 ; 0.2004]		
NE2 -> Guangdong -> Kasumig3	0,0679	[0.0352 ; 0.1006]		
NE2 -> Guangdong ; NE2+Guangdong -> Kasumig3	0,424	[0.3848 ; 0.4632]	0,75	0,14
Groupe de priors n°2 - 30 générations/an				
NE2 -> Kosa ; NE2 -> Laoshan	0.0923	[0.0632 ; 0.1214]		
NE2 -> Kosa -> Laoshan	0.9077	[0.8786 ; 0.9368]	0,03	0,06
NE2 -> Kosa ; NE2 -> Guangdong	0.8673	[0.8217 ; 0.9130]	0,04	0,02
NE2 -> Kosa -> Guangdong	0.1327	[0.0870 ; 0.1783]		
NE2 -> Kasumig3 ; NE2 -> Guangdong	0,3672	[0.3275 ; 0.4069]		
NE2 -> Kasumig3 -> Guangdong	0,2153	[0.1802 ; 0.2504]		
NE2 -> Guangdong -> Kasumig3	0,1041	[0.0766 ; 0.1317]		
NE2 -> Guangdong ; NE2+Guangdong -> Kasumig3	0,3134	[0.2760 ; 0.3508]		

Tableau 5 : Résultats des comparaisons de scénarii obtenus par la méthode ABC et impliquant toutes les zones envahies. NE2 = échantillon américain natif, Kosa = échantillon japonais, Laoshan = échantillon chinois, Guangdong = échantillon chinois et TR1 = échantillon portugais. Les probabilités *a posteriori* ont été obtenues par régression logistique avec l'option LDA (pour "Linear Analysis Discriminant", Estoup et al. 2012). Le scénario sélectionné sur la base des probabilités *a posteriori* et de leur intervalle de confiance est surligné en vert. Les erreurs de type I et de type II ont été calculé pour ce scénario.

Analyse et scénario	Probabilité <i>a posteriori</i>	Intervalle de confiance	Erreur type I	Erreur type II
Groupe de priors n°1				
30 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.6041	[0.5848 ; 0.6235]	0,12	0,013
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0103	[0.0000 ; 0.0428]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.1923	[0.1544 ; 0.2302]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.1933	[0.1652 ; 0.2215]		
15 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.5058	[0.4812 ; 0.5304]	0,11	0,0167
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0076	[0.0000 ; 0.0393]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.2443	[0.2209 ; 0.2677]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.2424	[0.2048 ; 0.2800]		
45 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.5115	[0.4926 ; 0.5305]	0,13	0,0267
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0144	[0.0000 ; 0.0369]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.2028	[0.1842 ; 0.2214]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.2713	[0.2443 ; 0.2983]		
Groupe de priors n°2				
30 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.6926	[0.6651 ; 0.7202]	0,12	0,0167
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0398	[0.0000 ; 0.0994]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.0781	[0.0178 ; 0.1383]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.1895	[0.1205 ; 0.2586]		
15 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.7658	[0.7378 ; 0.7938]	0,09	0,033
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0321	[0.0000 ; 0.1210]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.0595	[0.0000 ; 0.1492]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.1426	[0.0435 ; 0.2417]		
45 générations/an				
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong ; NE2 -> TR1	0.6442	[0.6197 ; 0.6686]	0,1	0,0133
NE2 -> Kosa -> Laoshan ; NE2 -> Guangdong -> TR1	0.0287	[0.0000 ; 0.0722]		
NE2 -> Kosa -> Laoshan et TR1 ; NE2 -> Guangdong	0.1025	[0.0604 ; 0.1447]		
NE2 -> Kosa -> Laoshan -> TR1 ; NE2 -> Guangdong	0.2246	[0.1743 ; 0.2748]		

❖ Discussion

L'objectif principal de cette étude était de clarifier les routes d'invasion du PWN en comparant différents scénarii d'invasion à l'aide de la méthode ABC. Le but ultime étant de déterminer l'origine des populations européennes.

En marge des routes d'invasion, le premier résultat obtenu dans cette étude est le taux de mutation faible que nous avons dû définir pour nos marqueurs microsatellite. En effet, nous avons dû changer l'intervalle de la distribution *a priori* du taux de mutation moyen et le faire passer de $10^{-4} - 10^{-3}$ (valeurs pas défaut) à $10^{-6} - 10^{-4}$ pour obtenir des distributions *a priori* de statistiques résumées simulées convenables vis-à-vis des statistiques résumées observées. Cela peut paraître surprenant dans la mesure où il est généralement admis que le taux de mutation moyen des marqueurs microsatellite est de l'ordre de 10^{-4} (Selkoe & Toonen 2006). Néanmoins, le taux de mutation des marqueurs microsatellite est possiblement dépendant de la taille du motif et du nombre de répétitions de ce motif (Schlötterer 2000; Ellegren 2004). Il a en effet été démontré dans différentes études que le taux de mutation pouvait effectivement varier en fonction de ces critères (voir Schug et al. 1997 pour un exemple chez la drosophile; et Molnar et al. 2012; Phillips et al. 2009 pour des exemples chez des nématodes). Les marqueurs utilisés dans cette étude sont presque exclusivement des dinucléotides et le nombre de répétitions des motifs n'excède pas 10 (Mallez et al. 2013), ce qui peut justifier l'intervalle de la distribution du taux de mutation défini ici et expliquer en partie la faible diversité rencontrée.

Concernant les routes d'invasion a proprement parlé, nous avons réalisé plusieurs comparaisons de scénarii de manière indépendante. Nous avons dans un premier temps étudié l'Asie, puis dans un second, l'Asie et l'Europe simultanément. L'ABC peut en théorie permettre de distinguer des scénarii très complexes, avec des histoires démographiques complexes, mais la quantité d'informations contenue dans les données est susceptible de

limiter la puissance de cette méthode pour les inférences précises (Stocks et al. 2014). Cette limite a été constatée lors de l'étude des relations entre Guangdong et Kasumig3. Néanmoins, la faible diversité génétique mise en évidence chez le PWN ne nous a pas empêché de sélectionner des scénarii (parmi ceux comparés) avec une probabilité *a posteriori* élevée et fiable dans la majorité des analyses effectuées. Nous avons ainsi pu déterminer globalement le scénario d'invasion mondiale du PWN le plus probable parmi ceux testés grâce aux différentes analyses effectuées.

Les analyses ABC conduites sur les échantillons asiatiques nous ont permis de démontrer l'existence d'au moins 3 introductions en provenance d'Amérique du Nord : une en Chine *via* le Japon, une 2^{ème} en Chine et une 3^{ème} en Europe. De plus, nous avons montré que le Japon avait connu au moins 2 introductions, dont une en provenance de l'aire native et l'autre ayant une origine que nous n'avons pas pu clairement déterminer. Des introductions multiples en Chine et au Japon sont donc plus probables que des introductions uniques, contrairement à ce que pouvait laisser penser la faible diversité génétique existante dans ces zones. Ce résultat a également été suggéré par l'analyse Bayésienne de groupement d'individus et l'arbre de distances génétiques. Particulièrement pour la Chine, ces résultats viennent appuyer l'hypothèse de deux événements d'introduction, le premier près de Nanjing en 1982 et le second près de Hong-Kong en 1988, présentée dans Robinet et al. (2009). Ils contredisent cependant l'hypothèse d'une introduction unique avancée par Cheng et al. (2008). Nous disposons cependant de peu d'échantillons chinois et japonais pour étudier finement l'histoire asiatique du PWN. De plus, la quantité d'information contenue dans nos données est aussi apparue limitante. Etudier plus d'échantillons provenant de ces 2 zones envahies avec un nombre de marqueurs plus important pourrait permettre de préciser les relations asiatiques et déterminer plus finement le scénario d'invasion asiatique.

Les analyses intégrant le Portugal ont mis en évidence qu'une origine américaine des populations européennes était plus probable qu'une origine asiatique, qu'elle soit japonaise ou chinoise. Ce résultat n'est pas apparu sensible aux changements de *priors* et s'est avéré robuste vis-à-vis des étapes de vérification réalisées. Les analyses ABC nous ont ainsi permis de trancher sur cette question, ce que les méthodes traditionnelles ne nous avaient pas permis de faire. L'origine américaine des populations portugaises était supportée par les valeurs de F_{st} et par la vraisemblance d'assignation individuelle alors que l'analyse Bayésienne de groupement des individus et l'arbre de distances génétiques avaient quant à eux suggéré une origine asiatique (voir partie III.1.b, Mallez et al. 2014). La non-concordance entre les résultats des méthodes traditionnelles et des analyses ABC obtenus sur ce cas d'invasion nécessite des analyses supplémentaires pour en expliquer l'origine. Nous avons abordé ce point dans la discussion générale (partie IV.2).

Ce résultat remet en question l'hypothèse d'une origine asiatique des populations portugaises, bien souvent avancée dans la littérature sur la base d'arbres de distances génétiques (Figueiredo et al. 2013; Fonseca et al. 2012; Metge & Burgermeister 2008; Valadas et al. 2012; Vieira et al. 2007). Avant le démarrage de ce travail, la majorité de la communauté "nématode du pin" était quasiment convaincue que l'Asie était la source des populations européennes, de sorte que ce résultat constitue une petite révolution. Il souligne que la génétique des populations, discipline assez peu utilisée dans le domaine de la nématologie (surtout sur des nématodes parasites de plantes), peut aider à comprendre et à améliorer nos connaissances sur l'histoire des nématodes par exemple, comme dans ce projet (Gilabert & Wasmuth 2013). D'un point de vue appliqué, ce résultat peut contribuer à prévenir une nouvelle introduction en Europe, en concentrant les efforts sur de meilleurs contrôles des importations américaines (contrôles qui se sont peut-être un peu relâchés du fait de l'idée dominante de l'origine asiatique des populations envahissantes portugaises).

4. Explorer les raisons du succès de l'invasion

Cette étude constitue une exploration préliminaire des facteurs qui ont pu conditionner le succès de l'invasion du PWN. Nous l'avons réalisée en comparant, au laboratoire, des populations naturelles originaires de différentes zones géographiques.

❖ Introduction

L'environnement d'introduction est potentiellement différent de l'environnement de l'aire native d'un point de vue biotique et abiotique. Les organismes introduits peuvent donc faire face à de nouvelles conditions environnementales et leur capacité à tolérer ou à s'adapter à ces nouvelles conditions influence le succès de leur établissement et de leur prolifération ultérieure (Facon et al. 2006). La sélection naturelle joue ainsi sûrement un rôle important et des phénomènes adaptatifs peuvent être à l'origine du succès des invasions (Lee 2002; Sakai et al. 2001; Wares et al. 2005). Ces phénomènes adaptatifs peuvent avoir lieu à différents moments au cours du processus d'invasion et différents scénarii sont possibles (Rey et al. 2012) : (i) adaptation après l'introduction (Lee 1999; Lombaert et al. 2010) ou (ii) adaptation dans l'aire native (Bossdorf et al. 2008; Foucaud et al. 2009; Hufbauer et al. 2012). Ignorer ces phénomènes peut conduire à des erreurs et mauvaises prédictions concernant le potentiel envahissant et l'impact écologique des organismes envahissants. Ainsi, déterminer ce qui a permis à un organisme (une population) de réussir son invasion est crucial pour améliorer notre connaissance des mécanismes de l'invasion et par conséquent les méthodes de contrôle.

Comparer les populations envahissantes et leurs populations sources dans l'aire native est un moyen de tester l'existence de phénomènes adaptatifs impliqués dans l'invasion. Cela permet en effet de mettre en évidence des différences phénotypiques éventuellement

adaptatives qu'il peut exister entre populations natives et envahissantes correspondantes, d'évaluer ce qui a pu évoluer au cours de l'invasion et d'expliquer son succès (Bossdorf et al. 2005; Keller & Taylor 2008).

Dans cette étude, nous avons procédé à de telles comparaisons avec notre modèle biologique, le PWN. Nous avons ainsi exploré par une expérience de laboratoire si des échantillons de populations native et envahissantes présentaient des différences en terme de dynamique de croissance.

❖ Matériel et méthodes

Échantillonnage des nématodes et préparation de l'expérience

Nous avons analysé 5 échantillons provenant de 4 zones géographiques différentes : USA (pour l'aire native), Japon, Chine et Portugal (pour les zones envahies, voir Tableau 6 pour le détail de chaque échantillon). Les différents échantillons ont été reçus sous forme de bois infesté et les nématodes vivants ont été extraits après passage sur tamis des morceaux de bois imprégnés d'eau pendant 24h. Une suspension de plusieurs centaines à quelques milliers de nématodes a ainsi été récupérée pour chaque échantillon de bois reçu. Quatre cent PWN nécessaires à l'expérience ont été récupérés un par un à partir de ces suspensions. Ils ont été différenciés des autres espèces de nématodes présents dans la suspension sur la base de critères morphologiques. En effet, la vérification selon des critères moléculaires nécessitant la mort du nématode, il nous était impossible de les appliquer sur les spécimens requis pour l'expérience. Néanmoins, pour chaque échantillon analysé, des nématodes supplémentaires ont été récupérés et vérifiés par analyses moléculaires. La majorité des nématodes, vérifiés *via* les analyses moléculaires, se sont avérés être des nématodes du pin, confirmant ainsi que notre tri sur critères morphologiques était bon. Les 400 nématodes du pin par échantillon

récupérés pour l'expérience ont permis de réaliser une étape de multiplication sur champignon (*Botrytis cinerea*) en boîtes de Pétri, ayant pour but d'obtenir un nombre suffisant de nématodes pour l'expérience et de minimiser les effets maternels et environnementaux. Deux boîtes de champignon ont été inoculées avec 200 nématodes chacune. Elles ont été incubées à 25°C pendant 3 semaines (soit environ 5 générations). Ces 2 boîtes de multiplication ont ensuite été utilisées pour lancer l'expérience proprement dite (présentée ci-dessous).

Il est important de souligner ici que tous les échantillons n'ont pas pu être traités simultanément parce qu'ils n'ont pas été reçus au même moment. Une souche de collection, Bx China, au laboratoire depuis 2008, a donc été utilisée comme souche témoin "fil rouge" et a été traitée en parallèle de chacun des 5 échantillons d'intérêt au cours de 5 manipulations déconnectées dans le temps (voir Tableau 6).

Tableau 6 : Présentation des échantillons comparés dans cette étude et de leur période d'expérience.

Nom de l'échantillon		Localisation	Période d'expérience
Aire Native	Boylston	USA - Mississippi	29 Octobre - 20 Novembre 2013
	Jap116	Japan - Iwate - Shiwa	29 Janvier - 20 Février 2013
Aire Envahie	Kasumig3	Japan - Ibaraki - Kasumigaura	28 Août - 19 Septembre 2013
	Laoshan	Chine - Shandong - Laoshan	8 Janvier - 30 Janvier 2014
	H.Comporta	Portugal - Setubal - Herdade de Comporta	24 Juin - 16 Juillet 2014
Laboratoire	BxChina	Chine - Guangdong	Toutes les périodes

Déroulement de l'expérience et comptages

L'expérience réalisée dans cette étude a consisté à inoculer des boîtes de champignon (*Botrytis cinerea*) avec une quantité connue de nématodes et à compter le nombre de nématodes obtenus au bout de 7, 14 et 21 jours. Nous avons ainsi inoculé trente boîtes de champignon pour les échantillons d'intérêt et 15 boîtes pour la souche témoin (Bx China) avec 100 nématodes chacune, issus de l'étape de multiplication. L'inoculation, qui nécessitait de récupérer, un par un, un grand nombre de nématodes, a limité le nombre de boîtes pouvant être traitées simultanément. Au bout de chacune des trois durées (7, 14 et 21 jours), nous avons compté les nématodes présents dans dix boîtes pour chaque échantillon d'intérêt et cinq

boîtes pour la souche témoin. Les nématodes ont été comptés de la manière suivante : scarification de la gélose des boîtes de Pétri avec un scalpel puis passage sur tamis pendant au moins 18h après imprégnation d'eau (voir Photo 1 ci-dessous). Nous avons ensuite récupéré une suspension de nématodes par boîte et avons effectué trois comptages par suspension. Le nombre total de nématodes par boîte a été obtenu en moyennant trois comptages de 1ml de suspension (réalisés avec des lames à compter) et en multipliant cette moyenne par le volume totale de la suspension récupérée.



Photo 3 : Passage de 10 boîtes passées sur tamis pour récupération des nématodes en vue des comptages.

Analyses statistiques

Toutes les analyses statistiques ont été réalisées grâce au logiciel R version 2.14.2 (R Development Core Team 2012).

Nous avons tout d'abord cherché à savoir si les 5 périodes d'expérience avaient eu un impact sur les comptages de la souche témoin à chaque durée (soit au bout de 7, 14 et 21 jours). Des tests de Kruskal-Wallis par durée ont donc été réalisés avec les données de comptage de la souche témoin. Des différences significatives de comptages de la souche témoin ont été constatées entre les 5 périodes d'expérience (voir Résultats). Nous avons donc tenu compte des comptages de la souche témoin pour analyser les comptages de chaque échantillon d'intérêt. Ainsi, pour chacune des 5 périodes d'expérience, nous avons divisé les

10 comptages obtenus à chaque durée pour chaque échantillon d'intérêt par la moyenne des comptages de la souche témoin. Ce ratio nous a permis de tenir compte d'un possible biais dû à la désynchronisation des expériences en 5 périodes. Il a constitué la variable à expliquer qui a été analysée dans les analyses présentées ci-dessous.

Les variations de ce ratio ont été analysées grâce à des modèles linéaires généralisés (GLM) en considérant le logarithme du ratio des comptages et une famille d'erreur gaussienne. Nous avons tout d'abord procédé à une comparaison de modèles sur la base du critère d'information d'Akaike (AIC) pour sélectionner le modèle expliquant le mieux nos données (soit les variations du ratio des comptages). Trois modèles ont été comparés :

(1) $\log(\text{ratio}) \sim \text{durée}$

(2) $\log(\text{ratio}) \sim \text{échantillons} + \text{durée}$

(3) $\log(\text{ratio}) \sim \text{échantillons} * \text{durée}$

La variable "durée" correspond aux différentes durées avant comptage (7, 14 et 21 jours) et a été considérée comme un facteur (à 3 modalités) car nous n'avions pas de raison de considérer un modèle linéaire pour la durée (bien qu'un modèle logistique aurait éventuellement pu être utilisé, mais aurait "consommé" autant de degrés de liberté qu'un facteur qualitatif). La variable "échantillons" correspond aux différents échantillons analysés dans cette expérience (5 modalités). La comparaison de modèles nous a conduits à sélectionner le modèle n°3, avec interaction (voir Résultats). Nous avons donc procédé à des analyses durée par durée. Ainsi, pour chaque durée, nous avons réalisé un GLM avec le logarithme du ratio des comptages comme variable à expliquer et la variable "échantillons" comme variable explicative. Nous avons ensuite utilisé la fonction "glht" du package R "multcomp" pour obtenir toutes les comparaisons d'échantillons grâce à la méthode post-hoc de Tukey (Bretz et al. 2010).

Ayant conscience que le choix d'une famille gaussienne pour la réalisation des GLM est critiquable, nous avons donc également procédé à des tests non-paramétriques pour analyser les variations du ratio calculé précédent. Dans un premier temps, nous avons réalisé un test de Kruskal-Wallis durée par durée pour chercher à savoir s'il existait des différences de ratio entre les différents échantillons. En cas d'un résultat significatif du test de Kruskal-Wallis, nous avons procédé à des comparaisons multiples entre paires d'échantillons. Nous avons pour cela utilisé la méthode de comparaisons multiples décrite dans Siegel & Castellan (1988 pages 213 – 214) et implémentée dans la fonction "kruskalmc" du package R "pgirmess". Dans un second temps, nous avons étudié plus précisément les différences qu'il pouvait exister entre l'échantillon natif, Boylston, et les échantillons des différentes zones envahies. Nous avons donc réalisé des tests de Wilcoxon entre Boylston et les différents échantillons des zones envahies pour déterminer quel(s) échantillon(s) envahissant(s) était(en)t différent(s) de l'échantillon natif. Pour ces dernières comparaisons, des tests multiples et non indépendants étant réalisés, le seuil de significativité a été ajusté par la méthode de Bonferroni séquentielle (Sokal & Rohlf 1995).

❖ Résultats

Dans certaines boîtes de Pétri, les nématodes inoculés ne se sont pas multipliés, ce qui a conduit à des valeurs de comptage aberrantes. Ces valeurs aberrantes (1 pour les boîtes de la souche témoin et 5 pour l'ensemble des 150 boîtes des échantillons d'intérêt) ont été supprimées du jeu de données dans la mesure où elles perturbaient les résultats (résultats non montrés). Un patron de croissance similaire a été obtenu pour les différentes répétitions de la souche témoin (correspondant aux 5 périodes d'expérience) et pour les échantillons d'intérêt. Ce patron présente deux phases : une phase de croissance, avec l'augmentation du nombre de nématodes à 7 et 14 jours et une phase de plateau, entre 14 et 21 jours (voir Figure 13).

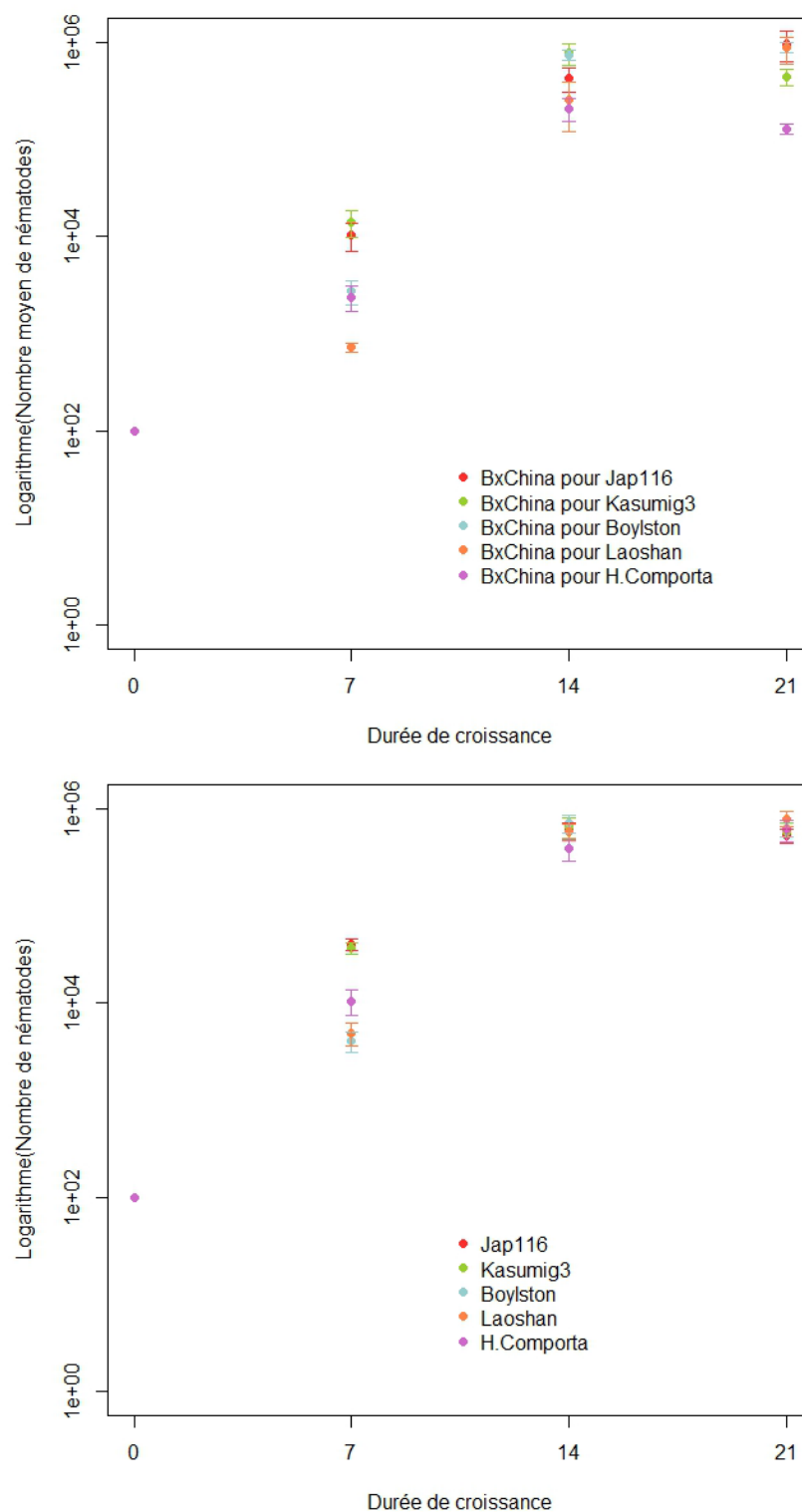


Figure 13 : Logarithme du nombre moyen de nématodes (\pm l'erreur standard) obtenus à 7, 14 et 21 jours après inoculation pour les différentes répétitions de la souche témoin, BxChina, en haut, et pour les échantillons d'intérêt en bas. La durée "0" correspond à l'inoculation (100 nématodes par boîte de Pétri).

Au sein de chaque durée considérée, les comptages de la souche témoin ne sont pas apparus homogènes entre les 5 périodes d'expérience (tests de Kruskal-Wallis à 7, 14 et 21 jours, $p = 0.0120$, 0.0123 et 0.0127 respectivement, voir Figure 13). Ce résultat a impliqué de tenir compte des comptages de la souche témoin pour analyser les comptages des échantillons d'intérêt. Nous l'avons fait en créant la variable "ratio" comme expliqué dans la partie "Analyses statistiques" du matériel et méthodes.

En ce qui concerne l'étude des variations de la variable "ratio", nous avons sélectionné le modèle n°3 [$\log(\text{ratio}) \sim \text{échantillons} * \text{durée}$] sur la base de l'AIC, impliquant que les différences entre échantillons variaient en fonction de la durée de croissance (AIC des modèles n°1, 2 et 3 : 412.06, 399.7 et 383.4, respectivement). Des GLM durée par durée ont donc été réalisés. La distribution de leurs résidus s'est avérée globalement satisfaisante, c'est-à-dire relativement proche d'une gaussienne. Les résultats des GLM sont représentés dans la figure 14. À 7 jours, aucune différence n'a été observée entre les différents échantillons des zones envahies (comparaisons multiples après le GLM, $p = 0.253$ à 0.999). Seuls deux échantillons des zones envahies, Jap116 et Laoshan, sont apparus significativement différents de l'échantillon natif, Boylston (comparaisons multiples après le GLM, $p = 0.02917$ et 0.00115 respectivement). À 14 jours, des différences significatives ont été observées entre Laoshan et Boylston (marginale) et Laoshan et Kasumig3 (comparaisons multiples après le GLM, $p = 0.05326$ et 0.00895 respectivement). Enfin, à 21 jours, l'échantillon portugais, H.Comporta, a présenté un logarithme du ratio positif significativement différent des ratios négatifs des autres échantillons (comparaisons multiples après le GLM, $p = 0.001$, 0.0191 , 0.001 et 0.006 pour les comparaisons avec Jap116, Kasumig3, Boylston et Laoshan, respectivement).

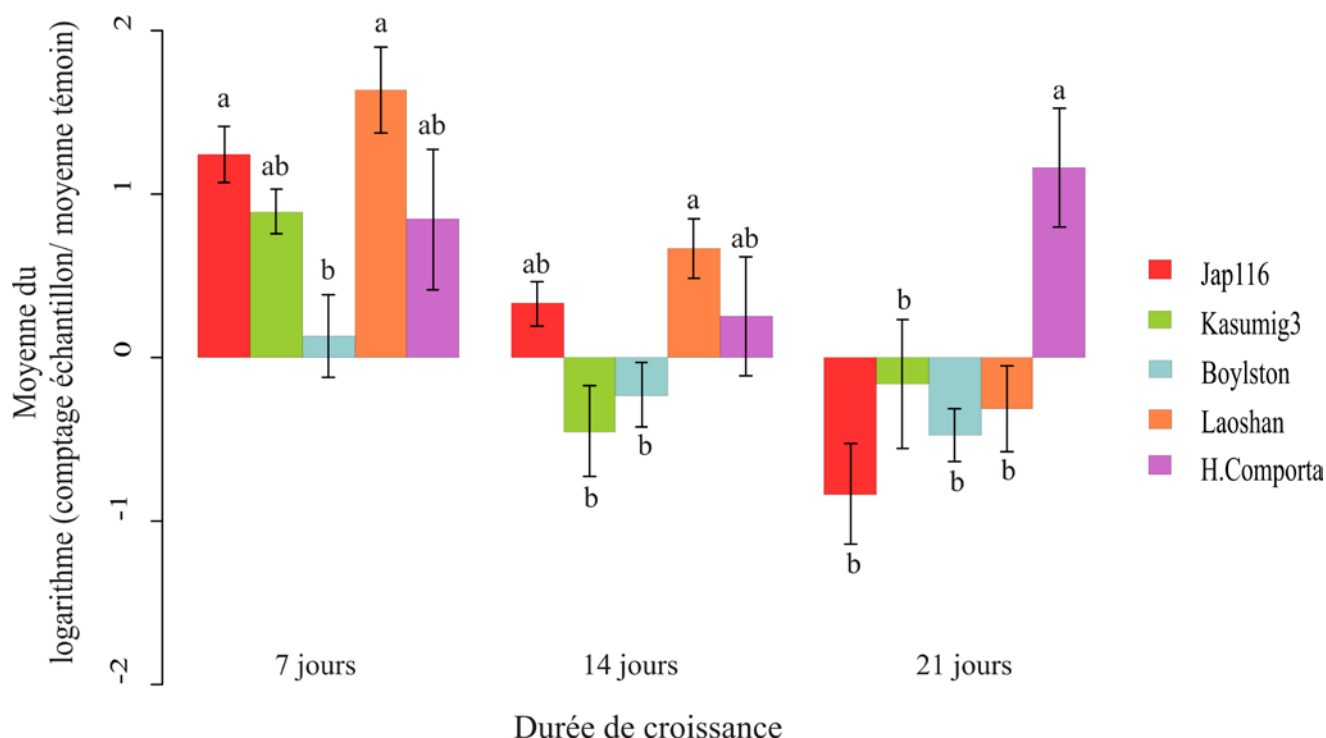


Figure 14 : Moyenne du logarithme du ratio obtenu en divisant les comptages des échantillons par la moyenne des comptages de la répétition de la souche témoin correspondante (\pm l'erreur standard). Les analyses ont été réalisées durée par durée et des lettres différentes au sein de chaque durée symbolisent des différences significatives (d'après les GLM réalisés). Boylston est l'échantillon natif et Jap116, Kasumig3, Laoshan et H.Comporta sont les échantillons des zones envahies.

Les résultats des tests non-paramétriques sont globalement concordants avec les GLM réalisés. Tous les tests de Kruskal-Wallis réalisés par durée se sont avérés significatifs, suggérant qu'il existait des différences significatives entre les différents échantillons ($p = 0.0084$, 0.0151 et 0.001 pour les tests réalisés à 7, 14 et 21 jours respectivement). Les comparaisons multiples réalisées ensuite ont permis de montrer que les résultats des tests de Kruskal-Wallis étaient dus à des différences significatives entre Boylston et Laoshan à 7 jours, Kasumig3 et Laoshan à 14 jours, et Boylston et H.Comporta, ainsi que Jap116 et H.Comporta à 21 jours. Les comparaisons entre Boylston et les différents échantillons des zones envahies ont de plus mis en évidence des différences significatives entre Boylston et Jap116 à 7 jours (test de Wilcoxon, $p = 0.0029$, significative après correction) et entre

Boylston et Laoshan à 14 jours (test de Wilcoxon, $p = 0.0052$, significative après correction). Ainsi, seules deux différences significatives à 21 jours observées avec les GLM, Kasumig3 - H.Comporta et Laoshan - H.Comporta, n'ont pas été mises en évidence avec les tests non-paramétriques.

❖ Discussion

Dans cette étude, nous avons comparé le nombre de nématodes obtenus au bout de 7, 14 et 21 jours de croissance, à partir d'un inoculum connu de 100 individus, pour différents échantillons de populations naturelles de PWN. Les différents échantillons ont été traités à des périodes différentes. Une souche témoin a donc été traitée en parallèle de chaque échantillon d'intérêt pour tenir compte du manque de synchronisation de notre expérience. Nous avons mis en évidence des différences significatives entre les 5 périodes d'expérimentation, ce qui a justifié l'utilisation de la souche témoin pour analyser les données de comptage obtenues pour les différents échantillons d'intérêt. Pour tenir compte des différences observées chez la souche témoin, nous avons créé la variable "ratio", obtenue en divisant les valeurs de comptages des différents échantillons d'intérêt par la moyenne des comptages de la répétition de la souche témoin correspondante et ce pour chaque durée. Cette variable nous a permis de comparer les différents échantillons d'intérêt en corrigeant le biais dû à la désynchronisation de l'expérience et donc de tenir compte des variations temporelles, induites par les différentes périodes de l'expérience. Nous sommes pleinement conscients que cette variable ne prend en revanche pas en compte des différences de variance qu'il pourraient y avoir entre les périodes d'expérimentation, différences qui pourraient avoir un impact sur les tests réalisés. Une réflexion approfondie sur ce point est nécessaire et sera menée.

L'analyse de cette variable a mis en évidence des différences significatives entre les échantillons, que ce soit entre les échantillons des zones envahies ou entre les échantillons des

zones envahies et l'échantillon natif. Deux échantillons de populations envahissantes, Laoshan et Jap116, présentent un ratio positif significativement plus grand que l'échantillon de la population native à 7 jours, soit pendant la phase de croissance. Bien que les résultats ne soient pas apparus significatifs, la même tendance a été retrouvée pour les autres échantillons envahissants. Ce résultat suggère que les échantillons envahissants tendent à présenter une croissance plus rapide que l'échantillon natif. Également, à 21 jours, seul l'échantillon envahissant H.Comporta présente un ratio positif et significativement différent de ceux des autres échantillons. Il apparaît donc que, dans un environnement similaire, pendant la phase de plateau de la croissance, H.Comporta peut constituer une population de plus grande taille que les autres échantillons, suggérant une meilleure capacité compétitive. Ainsi, les différences observées entre échantillons mettent en évidence l'existence de variabilité génétique sur des traits liées à la vitesse de croissance et peut-être aux capacités compétitives (qui doivent être explicitement évaluées par une expérience dédiée). Elles suggèrent donc que les populations envahissantes ont pu subir des changements évolutifs expliquant la réussite de l'invasion du PWN ou que l'invasion s'est faite à partir d'une population native qui a des caractéristiques génétiques différentes de celles de Boylston. Nous ne pouvons pas trancher entre ces deux hypothèses mais des améliorations du dispositif expérimental peuvent être apportées, qui pourraient fournir des informations complémentaires pertinentes.

La possibilité de recevoir des échantillons de bois infesté a conditionné les échantillons analysés. Ils n'ont donc pas été choisis de manière optimale, surtout au vu des résultats concernant les routes d'invasion (voir partie III.3), ce qui représente une limite importante de cette expérience. En effet, la comparaison entre populations sources et populations envahissantes nécessite une bonne connaissance des routes d'invasion pour ne pas conduire à des résultats erronés (Estoup & Guillemaud 2010). Cette expérience pourrait donc être améliorée en choisissant notamment des échantillons natifs plus appropriés. En effet, le

seul échantillon de l'aire native étudié ici, Boylston, s'avère très différencié génétiquement des autres échantillons de l'aire native et des échantillons envahissants que nous avons étudiés lors de ce projet. Cet échantillon n'est donc ni un bon représentant de l'aire native ni la source la plus probable des échantillons envahissants. De plus, au vu de la variation inter-échantillons et du niveau de structuration existant dans l'aire native, analyser plusieurs échantillons natifs apparaît nécessaire pour avoir une vision claire des différences phénotypiques qu'il peut exister entre populations natives et entre populations natives et envahissantes. De même, dans la mesure où de la variabilité sur des traits d'histoire de vie a été mise en évidence ici dans les zones envahies, analyser plusieurs échantillons par zone envahie pourrait aussi permettre d'affiner les résultats. La mise en évidence d'événements d'introductions multiples au Japon et en Chine justifie aussi la nécessité d'analyser différents échantillons de zones envahies.

En outre, l'échantillon natif que nous avons utilisé représente une population naturelle native issue d'un pin exotique sensible. Bien que cela soit logistiquement plus compliqué à obtenir (du fait de l'absence de symptômes dans le cas des pins natifs infestés), analyser des échantillons de populations naturelles natives présentes sur les arbres natifs serait utile pour préciser les raisons du succès de l'invasion du PWN. Il est en effet possible que le passage du PWN d'arbres natifs, qui ne souffrent pas de la PWD, aux arbres exotiques sensibles dans l'aire native ait pré-adapté le nématode à envahir en dehors de son aire native (comme suggéré par l'hypothèse de l'AIAI présenté dans Hufbauer et al. 2012). Comparer les capacités des trois catégories d'échantillons - échantillons de populations naturelles natives d'arbres natifs, de populations naturelles natives d'arbres sensibles et de populations envahissantes - pourrait permettre d'avoir une vision claire des différences qu'il peut exister entre ces trois catégories et donc de déterminer l'histoire ayant conduit le PWN à réussir son invasion.

IV. DISCUSSION GÉNÉRALE

Dans ce projet, nous avons étudié un cas précis d'invasion biologique, celui du PWN. L'objectif principal, au-delà de développer les outils nécessaires à la réalisation de ce projet, était d'analyser la diversité génétique des populations mondiales de PWN en lien avec l'invasion. Il s'agissait notamment de déterminer *in fine* les routes d'invasion du PWN, et plus précisément l'origine des populations européennes et le nombre d'évènements d'introduction qui ont eu lieu en Europe.

Dans la première partie de cette discussion générale, nous revenons sur les principaux résultats obtenus et présentons les limites et les perspectives de nos travaux. Dans la seconde, nous discutons de certains aspects des méthodes utilisées, qui nous ont interpellés au cours de cette thèse.

1. Invasion du nématode du pin

a. Principaux résultats

❖ Fonctionnement des populations

Avant de s'atteler à l'identification des routes d'invasion, nous nous sommes intéressés à l'aire native et avons tenté de déterminer le fonctionnement des populations naturelles natives de PWN, ce qui est utile pour comprendre les caractéristiques d'une invasion. Nos analyses sur les échantillons américains ont mis en évidence une diversité génétique faible à modérée dans une partie de l'aire native et un niveau de diversité génétique variable entre les échantillons étudiés. De plus, une structure génétique très forte et à fines échelles a également

été soulignée par les tests de différenciation génotypique et les analyses Bayésiennes de groupement d'individus effectués. Il s'avère donc que les populations naturelles natives de PWN souffrent de dérive génétique importante, qui n'est pas compensée par une dispersion efficace par l'insecte vecteur. Les capacités de dispersion supposées limitées de l'insecte vecteur peuvent expliquer ces résultats (Shibata 1986; Togashi 1990, pour des exemples avec le vecteur japonais). Également, le PWN fait partie d'un système biologique complexe comme expliqué en partie II.1 et doit s'affranchir de différents obstacles pour se reproduire dans l'arbre, ce qui peut, au-delà des capacités de dispersion de l'insecte, limiter sa dispersion (Kobayashi et al. 1984; Linit 1988; Togashi 1985).

❖ Introduction et routes d'invasion

Les différentes zones envahies étudiées ici, le Japon, la Chine et le Portugal/Madère, ont toutes présenté une diversité génétique faible, voire extrêmement faible, avec de nombreux allèles fixés. Ce résultat peut être expliqué par la faible diversité intra - échantillon dans l'aire native et par l'introduction d'un petit nombre d'individus. Bien que la relation diversité génétique neutre et capacités d'adaptation ne soit pas évidente (Dlugosch & Parker 2008; Dlugosch & Parker 2008; Reed & Frankham 2001), cette faible diversité génétique peut laisser penser que les capacités d'adaptation des populations envahissantes sont limitées (Reed & Frankham 2003; Willi et al. 2006). Ceci peut être un atout à considérer lors du développement de méthodes de lutte. Par exemple, il est légitime de supposer que le PWN aura des capacités réduites pour développer de la résistance contre des agents ou substances biologiques employés contre lui (comme des champignons, Wang et al. 2011, ou produits naturels dérivés des plantes, Faria et al. 2013) ou pour contrecarrer les défenses d'arbres résistants, suggérant que les efforts doivent être dirigés vers le contrôle biologique et/ou la création d'arbres résistants. Cette faible diversité génétique nous a également laissé supposer

qu'un seul évènement d'introduction avait eu lieu dans les différentes zones envahies. Précisément, en Europe, les populations de l'île de Madère et les populations du centre du Portugal ont présenté les mêmes allèles fixés que ceux observés au Portugal et près du point d'introduction respectivement, suggérant fortement un seul évènement d'introduction au Portugal, puis une expansion vers le centre du pays, ainsi qu'une introduction du Portugal continental vers l'île de Madère.

Les méthodes traditionnelles ne sont pas apparues concluantes pour déterminer l'origine des populations européennes. En effet, celles-ci ont présenté des résultats non-concordants, ce qui a justifié l'utilisation de la méthode ABC pour clarifier l'invasion du PWN. Au moins 3 introductions en provenance d'Amérique du Nord ont eu lieu : deux en Chine dont une *via* le Japon et une 3^{ème} au Portugal. De plus, deux introductions indépendantes ont eu lieu au Japon, dont au moins une en provenance de l'aire native. L'histoire de l'invasion asiatique s'est donc avérée bien plus complexe que la faible diversité génétique nous l'avait laissé supposer et des évènements d'introductions multiples sont apparus plus probables que des évènements d'introduction unique au Japon et en Chine. Le faible nombre d'échantillons asiatiques et la faible variabilité génétique des marqueurs utilisés ont limité la précision qu'il nous était possible d'atteindre dans cette étude. Des analyses supplémentaires sont donc nécessaires pour affiner ces résultats. Néanmoins, pour ce qui nous intéressait principalement dans ce projet, c'est-à-dire l'origine des populations européennes, l'ABC nous a permis de désigner l'Amérique du Nord comme la source la plus probable et d'exclure le Japon et la Chine comme sources potentielles.

Les analyses réalisées ici nous ont ainsi permis de clarifier en partie les routes d'invasion du PWN et particulièrement l'histoire de l'invasion européenne. Un évènement d'introduction unique au Portugal en provenance de l'aire native, puis une introduction du Portugal continental vers l'île de Madère s'avère donc le scénario le plus probable d'après les

analyses effectuées. Ce résultat contredit l'origine asiatique des populations européennes suggérée dans plusieurs études (Figueiredo et al. 2013; Fonseca et al. 2012; Metge & Burgermeister 2008; Valadas et al. 2012; Vieira et al. 2007), ainsi que l'hypothèse d'introductions multiples au Portugal avancée par Valadas et al. (2012). La non-concordance entre les résultats des méthodes traditionnelles et des analyses ABC obtenus sur ce cas d'invasion nécessite des analyses supplémentaires pour l'expliquer. La portée de l'explication dépasse largement notre cas d'étude. Nous avons donc abordé ce point dans la deuxième partie de cette discussion générale (partie IV.2.b).

❖ Raisons du succès

L'identification des routes d'invasion permet de réaliser des comparaisons entre populations sources et populations envahissantes pour tenter de déterminer les facteurs qui ont conditionné le succès d'une invasion biologique. De manière très préliminaire, nous avons exploré cela chez le PWN. Nous avons ainsi comparé le nombre de nématodes obtenus à partir d'un inoculum connu au bout de 7, 14 et 21 jours pour des échantillons natif et envahissants. Cette étude a mis en évidence des différences significatives entre les échantillons analysés, de sorte que les échantillons envahissants tendaient notamment à présenter une croissance plus rapide que l'échantillon natif. Ces résultats suggèrent que les populations envahissantes du PWN ont peut-être subi des changements évolutifs, pouvant expliquer le succès de leur invasion. Néanmoins, comme évoqué partie III.4, une analyse plus approfondie, précisément une expérience intégrant plus d'échantillons natifs et envahissants ainsi que des échantillons natifs plus appropriés, est nécessaire pour avoir une vision claire des différences phénotypiques qu'il peut exister entre populations natives et entre populations natives et envahissantes.

Outre l'hypothèse de changements évolutifs dans les populations envahissantes qui nécessite d'être approfondie, deux autres hypothèses non-exclusives peuvent être avancées pour expliquer l'invasion du PWN. (i) La présence d'hôtes sensibles et d'insectes vecteurs dans les zones envahies (Akbulut & Stamps 2012; Evans et al. 1996), couplée à la supériorité compétitive du PWN vis-à-vis d'espèces natives proches, comme *Bursaphelenchus mucronatus* (Cheng et al. 2009; Vincent et al. 2008), peuvent avoir facilité l'invasion du PWN sans que des changements adaptatifs soient nécessaires. Cette hypothèse abordée dans Mallez et al. (2014) peut expliquer le succès de l'invasion en dépit d'un manque de diversité génétique. (ii) Il est enfin possible que le passage du PWN d'arbres natifs, qui ne souffrent pas de la PWD, aux arbres exotiques sensibles dans l'aire native ait pré-adapté le nématode à envahir en dehors de son aire native, comme suggéré par l'hypothèse de l'AIAI (Hufbauer et al. 2012) et évoqué partie III.4.

❖ Effet "collection"

Un peu en marge des objectifs initiaux de ce projet, nous nous sommes intéressés à ce que nous avons appelé l'effet "collection". Dès la mise au point des marqueurs microsatellites, nous avons constaté une différence en terme de diversité génétique entre les échantillons de populations naturelles que nous avons obtenus et les souches de collection que le laboratoire possédait. Dans la mesure où de nombreuses études sur le PWN sont réalisées à l'aide de souches de collection, nous avons jugé utile d'explorer plus précisément ces différences. Pour cela, nous avons comparé les niveaux de diversité génétique et de différenciation existants entre des échantillons de populations naturelles et des souches de collection de l'aire native. Nous avons également simulé *in silico* le maintien d'une souche au laboratoire pour évaluer l'effet du maintien en collection sur la diversité génétique de la souche.

L'analyse des données a révélé des différences significatives entre les échantillons de populations naturelles et les souches de collection. En ce qui concerne la diversité génétique, le sens des différences est apparu dépendant de l'âge de la souche de collection. Nous avons en effet mis en évidence que les échantillons de populations naturelles étaient significativement moins diversifiés génétiquement que les souches récentes de collection mais plus diversifiés que les vieilles souches de collection. Concernant le niveau de différenciation, les vieilles souches sont apparues génétiquement plus différenciées entre elles et avec les échantillons de populations naturelles et les souches récentes de collection, que ces dernières. Les simulations réalisées ont confirmé les résultats obtenus par l'analyse de données vis-à-vis de la diversité génétique. En effet, de manière tout à fait inattendue, nous avons observé que le maintien en collection d'un échantillon de populations naturelles permettait d'augmenter sa diversité. En revanche, la survenue d'accidents lors du maintien la diminuait. Ce résultat est surprenant. Il contredit les hypothèses que nous avons initialement envisagées sur l'effet que pouvait avoir le maintien en collection sur la diversité génétique et va à l'encontre de ce qui est couramment observé lorsque des comparaisons populations naturelles – souches de collection sont réalisées. Beaucoup d'études montrent en effet une baisse de la diversité génétique dans les souches de collection par rapport aux populations naturelles (Arias et al. 2005; Bech et al. 2010; Ciosi et al. 2014; Coe et al. 2009; Fiumera et al. 2000; Norris et al. 2001; Nowak et al. 2007; Stohler et al. 2004), ce qui s'explique très bien par le confinement des souches de collection et donc l'absence de flux de gènes, et par les effets de fondation lors de la mise en collection.

L'augmentation de la diversité peut néanmoins s'expliquer par la simplification du cycle de vie du PWN au laboratoire : de très grandes tailles efficaces sont ainsi atteintes, limitent la dérive et permettent une augmentation de la diversité par l'accumulation de

mutations. De plus, la taille réduite de l'environnement des nématodes au laboratoire (boîtes de Pétri), permet un meilleur brassage des nématodes, diminuant la force de la dérive.

Bien qu'il soit nécessaire de clarifier le sens des différences observées, cette étude illustre les différences significatives qui peuvent exister entre les échantillons de populations naturelles et les souches de collection. Les effets observés peuvent biaiser les interprétations des résultats obtenus au laboratoire et mener à des conclusions erronées. D'autant plus lorsque les caractéristiques de la souche de collection, comme son niveau de diversité originelle, le nombre de fondateurs utilisés pour l'initier ou encore son âge, ne sont pas précisément connues, comme cela peut être le cas chez le PWN. Finalement, cette étude démontre aussi l'importance et l'utilité des simulations pour appréhender l'effet "collection" chez le PWN. Les simulations nous ont permis de reproduire le maintien d'une souche de collection en contrôlant les paramètres démographiques et de tester des hypothèses (l'impact des accidents par exemple). Elles nous ont ainsi permis d'évaluer l'effet du maintien sur la diversité de manière dynamique, ce qui est rarement possible ou logistiquement compliqué et ce qui nous a aidé à proposer des hypothèses expliquant les résultats obtenus en analysant les données génétiques.

b. Limites et perspectives

Dans ce projet, nous avons rencontré plusieurs obstacles dans ce projet, qui ont quelque peu limité la portée de nos travaux et qui nous ont parfois empêchés de mener à bien certaines analyses qui auraient aidé à améliorer la connaissance que l'on a de ce cas d'invasion. Ces limites sont présentées ici.

Le premier obstacle que nous avons rencontré a été la difficulté d'obtenir des échantillons de populations naturelles et de façon criante des échantillons vivants de

populations naturelles. Ce sont des collaborations et l'envoi massif de courriels (qui sont souvent restés sans réponse) qui nous ont permis d'obtenir les échantillons analysés dans ce projet. Le statut d'espèce de quarantaine du PWN a compliqué et ralenti l'envoi des échantillons ou a nécessité l'envoi d'individus morts (autorisations non nécessaires dans ce cas), ce qui a bien évidemment limité l'utilisation qu'il pouvait en être fait. Nous n'avons pas procédé nous-mêmes à l'échantillonnage sur le terrain et donc nous n'avons pu ni contrôler ni adapter le design de celui-ci en fonction des questions posées. Pourtant, pouvoir réaliser notre propre échantillonnage, particulièrement dans l'aire native, aurait probablement contribué à améliorer la compréhension du fonctionnement des populations naturelles de PWN. En effet, cela aurait permis d'étudier plus finement l'échelle de structuration des populations, par exemple en analysant différents échantillons de différentes parties d'un même arbre pour explorer la structure intra-arbre. Dans le même ordre d'idée, réaliser un vrai échantillonnage hiérarchique (plusieurs groupes d'arbres dans plusieurs localités) aurait permis d'affiner notre compréhension de la structuration des populations et indirectement de mieux appréhender la dispersion du PWN. Couplé à des patrons paysagers différents - arbres relativement isolés *versus* forêts denses - un tel échantillonnage nous aurait permis d'approfondir ce point et de tester si la dispersion était fonction de l'environnement. Enfin, réaliser un échantillonnage temporel aurait permis d'estimer la taille efficace des populations naturelles, en étudiant les différences de diversité génétique au cours du temps. Information qu'il aurait été utile de connaître pour la confronter à la diversité intra-arbre observée, pour étudier l'effet "collection" chez le PWN, pour définir plus précisément les *priors* dans les analyses ABC et pour rendre compte de l'efficacité de la transmission *via* le vecteur.

Le deuxième obstacle réside dans la faible - voire l'absence de - diversité génétique que nous avons observée dans les populations naturelles de PWN, particulièrement dans les zones envahies. Au-delà d'avoir peut-être perturbé certaines des analyses effectuées (point

discuté dans la partie suivante et les parties IV.2.a et IV.2.b), cette faible diversité nous a empêché d'étudier l'expansion post-introduction des populations européennes du PWN, sujet que nous voulions initialement aborder. En effet, en plus de l'intérêt théorique d'améliorer les connaissances des processus de colonisation/dispersion, étudier l'expansion des populations envahissantes aurait notamment été utile d'un point de vue appliqué dans notre cas. Cela aurait permis d'évaluer les effets des moyens actuels de confinement de l'infestation du PWN, c'est-à-dire la coupe franche de pins autour des foyers, pour tenter de les améliorer. Nous aurions ainsi pu déterminer, par des analyses de groupement d'individus spatialement explicites (Chen et al. 2007; Guillot et al. 2005) ou en testant l'isolement par la résistance (McRae 2006; McRae & Beier 2007), si les discontinuités géographiques, créées par les coupes, avaient réellement un impact sur la structure génétique spatiale des populations de PWN, c'est-à-dire indirectement sur la dispersion de l'insecte vecteur. Nous avons enfin l'idée de réaliser des études conjointes de génétique des populations sur le PWN et son insecte vecteur au Portugal, qui n'ont pas non plus été possibles en raison de l'absence de diversité génétique.

Également et comme déjà abordé à différentes reprises tout au long de la présentation de nos travaux, des analyses approfondissant certains aspects de nos travaux sont nécessaires. Elles représentent autant de perspectives, que nous avons reprises ici.

Tout d'abord, l'analyse d'échantillons de populations naturelles natives d'arbres natifs est une perspective importante. Ces populations constituent probablement le vrai compartiment naturel natif et celui-ci n'a malheureusement pas été étudié dans ce projet. Il est en effet très difficile d'y accéder étant donné que les arbres natifs ne souffrent pas de la PWD et n'ont de ce fait pas de symptômes apparents. Pourtant, obtenir ces populations permettrait de tester l'hypothèse de l'AIAI (Hufbauer et al. 2012) et d'affiner et compléter les résultats de l'expérience exploratoire que nous avons réalisée sur la croissance du PWN. En effet, nous

pourrions ainsi déterminer si le PWN était pré-adapté à envahir, en comparant les performances d'échantillons de populations naturelles natives d'arbres natifs, natives d'arbres exotiques et envahissantes. En outre, des mesures phytosanitaires sont également prises dans l'aire native et quand des arbres exotiques infestés par le PWN sont détectés, ils sont détruits. Il serait donc important de savoir si les populations naturelles natives retrouvées sur arbres natifs participent et s'intègrent dans la dynamique du PWN. Par exemple, déterminer si ces populations constituent des populations sources, des réservoirs "infectieux", est indispensable pour mieux appréhender le PWN dans l'aire native.

Finalement, préciser et améliorer notre compréhension de l'effet "collection" chez le PWN représente une seconde perspective. La plupart des collections de PWN n'ont pas été mises en place dans le but d'utiliser les souches pour réaliser des analyses de génétique des populations. Leur création n'a donc pas pris en compte certains facteurs importants pour de telles analyses, tels que la diversité originelle des populations naturelles dont sont issus les échantillons mis en collection, le nombre de fondateurs nécessaires au maintien de cette diversité originelle... Bien que cela soit très probablement lourd à mettre en œuvre, réaliser un suivi expérimental de la diversité génétique d'une souche de collection sur le long terme pourrait apporter des informations utiles. Ce suivi permettrait d'évaluer concrètement la diversité au cours du maintien de manière dynamique et l'impact des caractéristiques de l'établissement de la souche sur la diversité, tout en vérifiant les résultats obtenus à partir des simulations. Cela serait utile pour améliorer le protocole de maintien en vue de limiter les changements génétiques vis-à-vis de la population originelle. En effet, s'il s'avère que les simulations effectuées ont fidèlement reproduit ce qui se passe réellement dans la collection, elles pourraient constituer un outil utile et pratique pour la gestion des souches.

2. Au-delà de ce cas d'invasion

Pendant les analyses réalisées tout au long de ce projet, certains résultats obtenus avec des méthodes "consacrées" en génétique des populations nous ont étonnés et interpellés. Nous avons déjà évoqué la faible diversité génétique de notre modèle d'étude pour tenter d'expliquer cela et il est vrai que cette faible diversité nous a un peu perturbés. Mais finalement, elle a également pu être bénéfique. En effet, même si les divergences rencontrées ici entre différentes mesures et méthodes sont peut-être inhérentes au faible niveau de diversité rencontré, elles soulèvent tout de même des interrogations. Loin de pouvoir apporter des réponses définitives à ces interrogations, nous pensons néanmoins qu'il est utile d'en discuter. Nous sommes donc revenus dans cette partie sur les mesures de différenciation (F_{st} et D de Jost) et sur la non-concordance entre les méthodes traditionnelles et la méthode ABC. De plus, indépendamment du niveau de diversité génétique et de ses conséquences, ce projet a regroupé deux disciplines qui se rencontrent peu, la nématologie et la génétique des populations. Nous avons discuté les avantages que cette rencontre peut apporter.

a. Mesures de différenciation

Mesurer le niveau de différenciation génétique entre populations est une analyse classique et centrale en génétique des populations. Des tests statistiques de différenciation génétique et génotypique existent (par exemple ceux implémentés dans le logiciel Genepop, Rousset 2008) et si les tests effectués mettent en évidence des différences significatives entre populations, les mesures de différenciation prennent tout leur sens. L'important est alors d'évaluer l'ampleur de la différenciation. En effet, les mesures de différenciation permettent d'évaluer le niveau de structure des populations, de mesurer le temps de divergence entre

populations, de rendre compte de l'intensité des flux de gènes entre populations ou encore de déterminer les populations génétiquement singulières dans un but de conservation. Également, dans le contexte des invasions biologiques, ces mesures font aussi partie des outils utilisés pour tenter d'identifier la source potentielle de populations envahissantes.

Dans ce projet, nous avons utilisé le F_{st} de Weir & Cockerham (1984) pour mesurer le niveau de différenciation entre nos échantillons. Nous avons obtenu des valeurs de F_{st} très élevées et avons supposé que la faible variabilité intra-échantillon avait artificiellement augmenté les valeurs mesurées. La variabilité intra-échantillon peut effectivement fortement impacter le F_{st} au point de pouvoir induire en erreur lorsque l'on tire des conclusions à partir de cette mesure (Balloux et al. 2000; Jost 2008). Le F_{st} , de par sa définition, est dépendant de la variabilité intra-échantillon (Balloux et al. 2000; Gerlach et al. 2010; Gregorious 2010; Hedrick 1999; Heller & Siegismund 2009; Jost 2008). Ainsi, deux échantillons très variables (hétérozygotie élevée) auront tendance à avoir une valeur de F_{st} faible même s'ils sont très différents (Hedrick 1999; Jost 2008). De plus, des échantillons quasiment identiques mais très peu variables (hétérozygotie très faible) auront tendance à avoir une valeur de F_{st} élevée, comme nous l'avons observé dans le cas du PWN (Alcala et al. 2014; Gregorious 2010). Il est donc important de garder à l'esprit cette dépendance lorsque l'on fonde des raisonnements en écologie sur les valeurs mesurées de F_{st} .

Le F_{st} reste la mesure de différenciation la plus utilisée en génétique des populations, alors que d'autres mesures ont été proposées pour pallier ses faiblesses (Hedrick 2005; Jost 2008). Parmi elles, le D de Jost (2008) que nous avons utilisé. Cette mesure est basée sur la composition génétique plutôt que sur l'hétérozygotie (ce qui est le cas pour le F_{st}) et repose sur un partitionnement complet de la diversité génétique en des composantes intra-population et inter-population indépendantes, de sorte que la mesure de différenciation qui en découle n'est pas influencée par la diversité intra-population. Le D et le F_{st} se comportent

différemment (Heller & Siegismund 2009) et les valeurs maximales de ces deux mesures peuvent correspondre à des patrons génétiques différents (Verity & Nichols 2014). Plusieurs chercheurs se sont intéressés aux différences entre ces deux mesures (Alcala et al. 2014; Gerlach et al. 2010; Jost 2009; Verity & Nichols 2014; Whitlock 2011). Le D mesure le niveau de différenciation basée sur la composition génétique et le Fst mesure le niveau de différenciation basée sur la fixation d'allèles (Alcala et al. 2014; Gerlach et al. 2010). Le D apparaît ainsi pour certains comme la mesure réelle de la différenciation tandis que le Fst permet d'estimer les paramètres démographiques et mutationnels à l'origine de la différenciation, lorsque le modèle théorique de populations en île est considéré (Gerlach et al. 2010; Heller & Siegismund 2009; Jost 2009; Verity & Nichols 2014). Whitlock (2011) reste sceptique quant à l'utilité du D car cette mesure n'est reliée à aucune théorie évolutive et il n'y a donc pas de preuves formelles que ce qu'elle mesure soit utile pour les études en biologie évolutive et de la conservation. Apportant des preuves de l'utilité du D dans leur étude comparative sous le modèle théorique de populations en île, Alcala et al. (2014) montrent que le Fst et le D détectent des signaux différents de différenciation génétique dépendamment du régime mutationnel considéré (faible, intermédiaire et fort) et sont sensibles à différents aspects de la diversité génétique entre populations. Ils soulignent leur complémentarité et la nécessité de considérer les deux mesures lorsque l'on souhaite estimer la différenciation génétique. Ils précisent aussi que le D, au même titre que le Fst, peut être utilisé pour estimer les paramètres démographiques et mutationnels. De même, Verity & Nichols (2014) montrent la complémentarité du D et du Fst pour estimer le vrai niveau de différenciation et de mutation lorsque le taux de mutation est élevé et que le Fst seul est insuffisant pour cela.

Cette partie n'a pas la prétention de trancher sur la question de quelle mesure utiliser. Le D apparaît comme une mesure de différenciation plus cohérente que le Fst dans le cas du PWN. Pour autant, il s'avère que l'utilisation des mesures de différenciation pour détecter la

source probable des populations portugaises a donné un résultat concordant entre le Fst et la méthode ABC, contrairement au D qui suggérait une origine différente.

b. Méthodes traditionnelles et méthode ABC

Les méthodes traditionnelles (voir partie I.3.b) sont couramment utilisées pour étudier les routes d'invasion et notamment pour identifier les populations sources des populations envahissantes (Ascunce et al. 2011; Ciosi et al. 2008; Facon et al. 2003; Kolbe et al. 2004; Papura et al. 2012; Perdereau et al. 2013 pour des exemples). La discordance entre les résultats des analyses ABC, des Fst et des vraisemblances d'assignation individuelle d'une part et ceux de l'analyse de groupement d'individus et de l'arbre de distances génétiques d'autre part, obtenue en Europe dans cette étude soulève quelques interrogations.

L'extrêmement faible diversité génétique du PWN peut avoir perturbé les analyses effectuées. Les méthodes traditionnelles sont probablement plus perturbées que la méthode ABC par la faible diversité. En effet, cette dernière limite la puissance discriminatrice des analyses ABC et peut donc conduire à ne pas pouvoir déterminer le scénario le plus probable parmi les scénarii en compétition. En revanche, la faible diversité ne biaise *a priori* pas les résultats de l'ABC. En ce qui concerne les méthodes traditionnelles, la faible diversité peut peut-être engendrer un patron de structuration et/ou des relations entre échantillons que nous n'interprétons pas correctement. Par exemple, si l'aire native est peu diversifiée de sorte qu'elle présente des allèles très fréquents, il est probable que deux introductions indépendantes à partir de cette aire native (native → env1 et native → env2) conduisent à des échantillons qui seront plus proches entre eux que de leur aire native. Ces résultats seront interprétés comme un cas d'introductions successives (native → env1 → env2). Les méthodes de groupement (Bayésien comme avec le logiciel Structure ou construction d'arbres) posent

en effet le problème de ne pas présenter de support statistique permettant d'objectiver le choix des scénarii évolutifs (critique présentée dans Estoup & Guillemaud 2010, et dans les parties I.3.b et III.1.b).

Nous comptons réaliser des simulations informatiques permettant d'évaluer l'effet de la diversité de l'aire native sur les méthodes de groupements et sur l'ABC. Plus précisément, des jeux de données seront simulés sous un scénario d'invasion déterminé (et donc connu) en faisant varier le niveau de diversité génétique. Les jeux de données obtenus seront ensuite analysés grâce aux méthodes de groupement et ABC comme s'il s'agissait de jeux de données observées. Cela permettra d'évaluer le comportement de ces méthodes vis-à-vis du niveau de diversité génétique considéré. Différents scénarii d'invasion pourront également être considérés de sorte qu'il nous sera possible d'évaluer les patrons de structuration et les relations entre échantillons obtenus sous tel ou tel scénario. Ces analyses seront ainsi utiles pour (i) déterminer l'impact qu'a réellement la faible diversité sur les résultats obtenus par ces méthodes et (ii) vérifier que les interprétations qui sont faites à partir de ces méthodes dans le cadre des invasions biologiques sont appropriées.

c. Les nématodes et la génétique des populations

La nématologie et la génétique des populations ne se rencontrent que très peu surtout lorsqu'il s'agit de nématodes parasites de plantes. Ces modèles animaux ne se prêtent en effet pas bien à la génétique des populations pour plusieurs raisons. Tout d'abord, de nombreuses espèces parasites de plantes sont parthénogénétiques, ce qui a pu limiter leur attrait pour des généticiens. En effet, le cadre théorique en génétique des populations pour étudier de telles espèces n'est pas aussi bien établi que pour les espèces sexuées et les méthodes couramment utilisés en génétique des populations ne sont pas applicables aux organismes clonaux (mais

voir Halkett et al. 2005 et Arnaud-Haond et al. 2007). Ensuite, la plupart d'entre eux sont microscopiques et vivent dans des endroits difficilement accessibles, notamment le sol. Cela rend complexe leur échantillonnage, surtout au niveau individuel, ainsi que leur observation en milieu naturel. Enfin, la petite quantité d'ADN obtenue à partir de chaque individu complique l'utilisation de marqueurs génétiques. Néanmoins, des méthodes permettant de traiter la clonalité existent (Halkett et al. 2005; Arnaud-Haond et al. 2007) et des progrès techniques concernant la recherche de marqueurs, leur amplification et leur séquençage ont été réalisés (Gilabert & Wasmuth 2013; Karaïskou & Primmer 2008), pouvant pallier la majorité de ces limites.

La génétique des populations permet d'étudier l'histoire et la structure des populations et peut de fait permettre d'éclaircir des aspects importants pour la compréhension et le contrôle des nématodes phytoparasites. Par exemple, elle permet de délimiter les populations, ce qui est utile (i) pour définir l'échelle géographique à laquelle appliquer les mesures de contrôle et (ii) pour établir les capacités de dispersion active et passive (vecteurs) du nématode (Picard & Plantard 2006). Elle permet également d'étudier le mode de reproduction des nématodes et de mettre en évidence l'existence d'évènements de reproduction sexuée par exemple, ce qui peut avoir un impact important sur leur virulence (Villate et al. 2010). Ce projet ainsi que d'autres études montrent enfin l'utilité de la génétique des populations pour retracer l'histoire des populations de nématodes, déterminer leur relation et leur origine, ce qui est notamment important pour prévenir leur expansion future et développer des stratégies de contrôle efficace (Boucher et al. 2013; Plantard et al. 2008).

Comme ces quelques exemples le montrent, de plus en plus d'études sont le fruit de collaborations entre les nématologistes et les généticiens des populations, à l'instar de ce projet de thèse, et toutes prouvent l'intérêt et l'utilité de ces collaborations (Gilabert & Wasmuth 2013).

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abelleira, A., A. Picoaga, J. P. Mansilla and O. Aguin (2011). "Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain." Plant Disease **95**(6): 776.
- Akbulut, S. and W. T. Stamps (2012). "Insect vectors of the pinewood nematode: a review of the biology and ecology of Monochamus species." Forest Pathology **42**(2): 89-99.
- Alcala, N., J. Goudet and S. Vuilleumier (2014). "On the transition of genetic differentiation from isolation to panmixia: What we can learn from Gst and D." Theoretical Population Biology **93**(0): 75-84.
- Allendorf, F. W. and L. L. Lundquist (2003). "Introduction: Population biology, evolution, and control of invasive species." Conservation Biology **17**(1): 24-30.
- Arias, L., E. E. Bejarano, E. Marquez, J. Moncada, I. Velez and S. Uribe (2005). "Mitochondrial DNA divergence between wild and laboratory populations of *Anopheles albimanus* Wiedemann (Diptera : Culicidae)." Neotropical Entomology **34**(3): 499-506.
- Arnaud-Haond, S., C. M. Duarte, F. Alberto and E. A. Serrão (2007). "Standardizing methods to address clonality in population studies." Molecular Ecology **16**: 5115 - 5139.
- Ascunce, M. S., C. C. Yang, J. Oakey, L. Calcaterra, W. J. Wu, C. J. Shih, J. Goudet, K. G. Ross and D. Shoemaker (2011). "Global Invasion History of the Fire Ant *Solenopsis invicta*." Science **331**(6020): 1066-1068.
- Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser and J. Goudet (2000). "Microsatellites can be misleading: an empirical and simulation study." Evolution **54**(4): 1414-1422.
- Barres, B., J. Carlier, M. Seguin, C. Fenouillet, C. Cilas and V. Ravigne (2012). "Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation." Heredity **109**(5): 269-279.
- Beaumont, M. A., R. Nielsen, C. Robert, J. Hey, O. Gaggiotti, L. Knowles, A. Estoup, M. Panchal, J. Corander, M. Hickerson, S. A. Sisson, N. Fagundes, L. Chikhi, P. Beerli, R. Vitalis, J.-M. Cornuet, J. Huelsenbeck, M. Foll, Z. Yang, F. Rousset, D. Balding and L. Excoffier (2010). "In defence of model-based inference in phylogeography." Molecular Ecology **19**(3): 436-446.
- Beaumont, M. A. and B. Rannala (2004). "The Bayesian revolution in genetics." Nature Reviews Genetics **5**(4): 251-261.
- Beaumont, M. A., W. Y. Zhang and D. J. Balding (2002). "Approximate Bayesian computation in population genetics." Genetics **162**(4): 2025-2035.
- Bech, N., S. Beltran, J. Portela, A. Rognon, J.-F. Allienne, J. Boissier and A. Théron (2010). "Follow-up of the genetic diversity and snail infectivity of a *Schistosoma mansoni* strain from field to laboratory." Infection, Genetics and Evolution **10**(7): 1039-1045.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste and F. Bonhomme (1996-2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier Montpellier (France).
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate - a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society Series B-Methodological **57**(1): 289-300.
- Bergdahl, D. R. (1988). "Impact of pinewood nematode in North America : Present and future." Journal of Nematology **20**(2): 260 - 265.

- Bertorelle, G., A. Benazzo and S. Mona (2010). "ABC as a flexible framework to estimate demography over space and time: some cons, many pros." Molecular Ecology **19**(13): 2609-2625.
- Bleeker, W. (2003). "Hybridization and *Rorippa austriaca* (Brassicaceae) invasion in Germany." Molecular Ecology **12**: 1831 - 1841.
- Boissin, E., B. Hurley, M. J. Wingfield, R. Vasaitis, J. Stenlid, C. Davis, P. de Groot, R. Ahumada, A. Carnegie, A. Goldarazena, P. Klasmer, B. Wermelinger and B. Slippers (2012). "Retracing the routes of introduction of invasive species: the case of the *Sirex noctilio* woodwasp." Molecular Ecology **21**(23): 5728-5744.
- Bossdorf, O., H. Auge, L. Lafuma, W. Rogers, E. Siemann and D. Prati (2005). "Phenotypic and genetic differentiation between native and introduced plant populations." Oecologia **144**(1): 1-11.
- Bossdorf, O., A. Lipowsky and D. Prati (2008). "Selection of preadapted populations allowed *Senecio inaequidens* to invade Central Europe." Diversity and Distributions **14**(4): 676-685.
- Boubou, A., A. Migeon, G. K. Roderick, P. Auger, J. M. Cornuet, S. Magalhaes and M. Navajas (2012). "Test of Colonisation Scenarios Reveals Complex Invasion History of the Red Tomato Spider Mite *Tetranychus evansi*." Plos One **7**(4).
- Boucher, A. C., B. Mimee, J. Montarry, S. Bardou-Valette, G. Belair, P. Moffett and E. Grenier (2013). "Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and determination of the origin of populations in Quebec, Canada." Molecular Phylogenetics and Evolution **69**(1): 75-82.
- Boyd, I. L., P. H. Freer-Smith, C. A. Gilligan and H. C. J. Godfray (2013). "The consequence of tree pests and diseases for ecosystem services." Science **342**(6160).
- Bretz, F., T. Hothorn and P. Westfall (2010). "Multiple comparisons using R". Edited by Chapman & Hall/CRC.
- Cardoso, J. M. S., L. Fonseca and I. Abrantes (2012). "Direct molecular detection of the pinewood nematode, *Bursaphelenchus xylophilus*, from pine wood, bark and insect vector." European Journal of Plant Pathology **133**(2): 419-425.
- Castagnone, C., P. Abad and P. Castagnone-Sereno (2005). "Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda : Aphelenchoididae)." European Journal of Plant Pathology **112**(2): 191-193.
- Cavalli-Sforza, L. L. and A. W. F. Edwards (1967). "Phylogenetic analysis. Models and estimation procedures." The American Journal of Human Genetics **19**: 233 - 257.
- Chakraborty, R. and L. Jin (1993). "A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances." EXS **67**: 153-175.
- Chapple, D. G., K. A. Miller, F. Kraus and M. B. Thompson (2013). "Divergent introduction histories among invasive populations of the delicate skink (*Lampropholis delicata*): has the importance of genetic admixture in the success of biological invasions been overemphasized?" Diversity and Distributions **19**: 134 - 146.
- Chapuis, M.-P. and A. Estoup (2007). "Microsatellite null alleles and estimation of population differentiation." Molecular Biology and Evolution **24**(3): 621-631.
- Chen, C., E. Durand, F. Forbes and O. François (2007). "Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study." Molecular Ecology Notes **7**: 747 - 756.
- Chen, Z. (2011). "Is the weighted z-test the best method for combining probabilities from independent tests?" Journal of Evolutionary Biology **24**(4): 926-930.
- Cheng, X. Y., F. X. Cheng, R. M. Xu and B. Y. Xie (2008). "Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: Aphelenchoididae) and its possible spread routes in China." Heredity **100**(4): 356-365.
- Cheng, X. Y., P. Z. Xie, F. X. Cheng, R. M. Xu and B. Y. Xie (2009). "Competitive displacement of the native species *Bursaphelenchus mucronatus* by an alien species *Bursaphelenchus*

- xylophilus (Nematoda: Aphelenchida: Aphelenchoididae): a case of successful invasion." Biological Invasions **11**: 205 - 213.
- Ciosi, M., D. K. Masiga and C. M. R. Turner (2014). "Laboratory colonisation and genetic bottlenecks in the tsetse fly *Glossina pallidipes*." PLoS Negl Trop Dis **8**(2): e2697.
- Ciosi, M., N. J. Miller, K. S. Kim, R. Giordano, A. Estoup and T. Guillemaud (2008). "Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity." Molecular Ecology **17**(16): 3614-3627.
- Coe, T. S., P. B. Hamilton, A. M. Griffiths, D. J. Hodgson, M. A. Wahab and C. R. Tyler (2009). "Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies." Ecotoxicology **18**(1): 144-150.
- Cohen, A. N. and J. T. Carlton (1998). "Accelerating invasion rate in a highly invaded estuary." Science **279**: 555 - 558.
- Commission Européenne (2006). "Décision de la commission du 13 février 2006." Journal officiel de l'Union européenne **52**: 34 - 38.
- Commission Européenne (2009). "Décision de la commission du 28 mai 2009." Journal officiel de l'Union européenne **135**: 29 - 32.
- Cornuet, J. M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J. M. Marin and A. Estoup (2014). "DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data." Bioinformatics **30**(8): 1187-1189.
- Cornuet, J. M., V. Ravigne and A. Estoup (2010). "Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0)." Bmc Bioinformatics **11**.
- Cornuet, J. M., F. Santos, M. A. Beaumont, C. P. Robert, J. M. Marin, D. J. Balding, T. Guillemaud and A. Estoup (2008). "Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation." Bioinformatics **24**(23): 2713-2719.
- Crawford, N. G. (2010). "SMOGLD: software for the measurement of genetic diversity." Molecular Ecology Resources **10**(3): 556-557.
- Csillery, K., M. G. B. Blum, O. E. Gaggiotti and O. Francois (2010). "Approximate Bayesian Computation (ABC) in practice." Trends in Ecology & Evolution **25**(7): 410-418.
- Daehler, C. C. (2001). "Two ways to be an invader, but one is more suitable for Ecology." Bulletin of the Ecological Society of America **82**(1): 101-102.
- Davis, M. and K. Thompson (2000). "Eight ways to be a colonizer ; Two ways to be an invader: a proposed nomenclature scheme for invasion ecology." Bulletin of the Ecological Society of America **81**(3): 226-230.
- Dlugosch, K. M. and I. M. Parker (2008). "Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions." Molecular Ecology **17**(1): 431-449.
- Dlugosch, K. M. and I. M. Parker (2008). "Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks." Ecology Letters **11**(7): 701-709.
- Drake, J. M. (2006). "Heterosis, the catapult effect and establishment success of a colonizing bird." Biology Letters **2** 304-307.
- Dropkin, V. H., A. Foudin, E. Kondo, M. J. Linit, M. Smith and K. Robbins (1981). "Pinewood nematode : a threat to US forests?" Plant Disease **65**(12): 1022 - 1027.
- Dwinell, L. D. (1985). "Relative susceptibilities of five pine species to three populations of the pinewood nematode." Plant Disease **69**(5): 440 - 442.
- Dwinell, L. D. (1997). "The pinewood nematode: Regulation and mitigation." Annual Review of Phytopathology **35**: 153-166.
- Dwinell, L. D. and W. R. Nickle (1989). An Overview of the Pine Wood Nematode Ban in North America. General Technical Report SE-55. N. A. F. C. Publication, United States Department of Agriculture - Forest Service: 1 - 20.

- Earl, D. and B. vonHoldt (2012). "STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method." Conservation Genetics Resources **4**(2): 359-361.
- Ellegren, H. (2004). "Microsatellites: simple sequences with complex evolution." Nat Rev Genet **5**(6): 435-445.
- Ellstrand, N. C. and K. A. Schierenbeck (2000). "Hybridization as a stimulus for the evolution of invasiveness in plants?" Proceedings of the National Academy of Sciences of the United States of America **97**(13): 7043-7050.
- Enserink, M. (1999). "Predicting invasions: Biological invaders sweep in." Science **285**(5435): 1834-1836.
- Estoup, A. and T. Guillemaud (2010). "Reconstructing routes of invasion using genetic data: why, how and so what?" Molecular Ecology **19**(19): 4113-4130.
- Estoup, A., E. Lombaert, J.-M. Marin, T. Guillemaud, P. Pudlo, C. P. Robert and J.-M. Cornuet (2012). "Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics." Molecular Ecology Resources **12**(5): 846-855.
- Evanno, G., S. Regnaut and J. Goudet (2005). "Detecting the number of clusters of individuals using the software structure: a simulation study." Molecular Ecology **14**(8): 2611-2620.
- Evans, H. F., D. G. McNamara, H. Braasch, J. Chadoeuf and C. Magnusson (1996). "Pest Risk Analysis (PRA) for the territories of European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*." OEPP/EPPO Bulletin **26**(2): 199 - 249.
- Facon, B., B. J. Genton, J. Shykoff, P. Jarne, A. Estoup and P. David (2006). "A general eco-evolutionary framework for understanding bioinvasions." Trends in Ecology & Evolution **21**(3): 130-135.
- Facon, B., R. A. Hufbauer, A. Tayeh, A. Loiseau, E. Lombaert, R. Vitalis, T. Guillemaud, J. G. Lundgren and A. Estoup (2011). "Inbreeding Depression Is Purged in the Invasive Insect *Harmonia axyridis*." Current Biology **21**(5): 424-427.
- Facon, B., P. Jarne, J. P. Pointier and P. David (2005). "Hybridization and invasiveness in the freshwater snail *Melanoides tuberculata*: hybrid vigour is more important than increase in genetic variance." Journal of Evolutionary Biology **18**(3): 524-535.
- Facon, B., J. P. Pointier, M. Glaubrecht, C. Poux, P. Jarne and P. David (2003). "A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails." Molecular Ecology **12**(11): 3027-3039.
- Facon, B., J. P. Pointier, P. Jarne, V. Sarda and P. David (2008). "High genetic variance in life-history strategies within invasive populations by way of multiple introductions." Current Biology **18**(5): 363-367.
- Faria, J. M. S., P. Barbosa, R. N. Bennett, M. Mota and A. C. Figueiredo (2013). "Bioactivity against *Bursaphelenchus xylophilus*: Nematotoxics from essential oils, essential oils fractions and decoction waters." Phytochemistry **94**: 220-228.
- Fielding, N. J. and H. F. Evans (1996). "The pine wood nematode *Bursaphelenchus xylophilus* (Steiner and Buhner) Nickle (= *B. lignicolus* Mamiya and Kiyohara) : an assessment of the current position." Forestry **69**(1): 35 - 46.
- Figueiredo, J., M. J. Simões, P. Gomes, C. Barroso, D. Pinho, L. Conceição, L. Fonseca, I. Abrantes, M. Pinheiro and C. Egas (2013). "Assessment of the Geographic Origins of Pinewood Nematode Isolates via Single Nucleotide Polymorphism in Effector Genes." Plos One **8**(12): e83542.
- Fiumera, A. C., P. G. Parker and P. A. Fuerst (2000). "Effective Population Size and Maintenance of Genetic Diversity in Captive-Bred Populations of a Lake Victoria Cichlid
- Tamaño Poblacional Efectivo y Mantenimiento de la Diversidad Genética en Poblaciones de un Cíclido del Lago Victoria Reproducidas en Cautiverio." Conservation Biology **14**(3): 886-892.

- Fonseca, L., J. M. S. Cardoso, A. Lopes, M. Pestana, F. Abreu, N. Nunes, M. Mota and I. Abrantes (2012). "The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island." Helminthologia **49**(2): 96 - 103.
- Fontaine, M. C., P. Gladieux, M. E. Hood and T. Giraud (2013). "History of the invasion of the anther smut pathogen on *Silene latifolia* in North America." New Phytologist **198**(3): 946-956.
- Foucaud, J., J. Orivel, D. Fournier, J. H. C. Delabie, A. Loiseau, J. Le Breton, P. Cerdan and A. Estoup (2009). "Reproductive system, social organization, human disturbance and ecological dominance in native populations of the little fire ant, *Wasmannia auropunctata*." Molecular Ecology **18**(24): 5059-5073.
- Frankham, R. (2005). "Resolving the genetic paradox in invasive species." Heredity **94**(4): 385-385.
- Futai, K. (2013). "Pinewood nematode, *Bursaphelenchus xylophilus*." Annual Review of Phytopathology **51**: 61-83.
- Garza, J. C. and E. G. Williamson (2001). "Detection of reduction in population size using data from microsatellite loci." Molecular Ecology **10**(2): 305-318.
- Gerlach, G., A. Jueterbock, P. Kraemer, J. Deppermann and P. Harmand (2010). "Calculations of population differentiation based on *Gst* and *D*: forget *Gst* but not all of statistics!" Molecular Ecology **19**(18): 3845-3852.
- Gilabert, A. and J. D. Wasmuth (2013). "Unravelling parasitic nematode natural history using population genetics." Trends in Parasitology **29**(9): 438-448.
- Golani, D., E. Azzurro, M. Corsini-Foka, M. Falautano, F. Andaloro and G. Bernardi (2007). "Genetic bottlenecks and successful biological invasions: the case of a recent Lessepsian migrant." Biology Letters **3**(5): 541-545.
- Goldstein, D. B., A. R. Linares, L. L. Cavallisforza and M. W. Feldman (1995). "An evaluation of genetic distances for use with microsatellite loci." Genetics **139**(1): 463-471.
- Goudet, J. (2002). "Fstat, a program to estimate and test gene diversities and fixation indices." from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Gregorious, H.-R. (2010). "Linking diversity and differentiation." Diversity **2**: 370 - 394.
- Grinnell, J. (1919). "The English Sparrow Has Arrived in Death Valley: An Experiment in Nature." The American Naturalist **53**(628): 468-472.
- Guillemaud, T., M. Ciosi, E. Lombaert and A. Estoup (2011). "Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics." Comptes Rendus Biologies **334**(3): 237-246.
- Guillot, G., F. Mortier and A. Estoup (2005). "GENELAND : a computer package for landscape genetics." Molecular Ecology Notes **5**: 712 - 715.
- Halkett, F., J.-C. Simon, et al. (2005). "Tackling the population genetics of clonal and partially clonal organisms." Trends in Ecology & Evolution **20**(4): 194-201.
- Hamilton, M. B. (2009). "Population Genetics". Edited by Wiley - Blackwell.
- Hedrick, P. W. (1999). "Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation." Evolution **53**(2): 313-318.
- Hedrick, P. W. (2005). "A standardized genetic differentiation measure." Evolution **59**(8): 1633-1638.
- Heller, R. and H. R. Siegmund (2009). "Relationship between three measures of genetic differentiation *GST*, *DEST* and *G'ST*: how wrong have we been?" Molecular Ecology **18**(10): 2080-2083.
- Hufbauer, R. A., B. Facon, V. Ravigné, J. Turgeon, J. Foucaud, C. E. Lee, O. Rey and A. Estoup (2012). "Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions." Evolutionary Applications **5**(1): 89-101.
- Hugot, J. P., P. Baujard and S. Morand (2001). "Biodiversity in helminths and nematodes as a field of study: an overview." Nematology **3**(3): 199 - 208.

- Jakobsson, M. and N. A. Rosenberg (2007). "CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure." Bioinformatics **23**(14): 1801 - 1806.
- Jombart, T., S. Devillard and F. Balloux (2010). "Discriminant analysis of principal components: a new method for the analysis of genetically structured populations." BMC Genetics **11**: 94.
- Jones, J. T., M. Moens, M. Mota, H. M. Li and T. Kikuchi (2008). "*Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host-parasite interactions." Molecular Plant Pathology **9**(3): 357-368.
- Jost, L. (2008). "G(ST) and its relatives do not measure differentiation." Molecular Ecology **17**(18): 4015-4026.
- Jost, L. (2009). "D vs. GST: Response to Heller and Siegmund (2009) and Ryman and Leimar (2009)." Molecular Ecology **18**(10): 2088-2091.
- Jung, J., H. Han, S. H. Ryu and W. Kim (2010). "Amplified fragment length polymorphism analysis and genetic variation of the pinewood nematode *Bursaphelenchus xylophilus* in South Korea." Animal Cells and Systems **14**(1): 31-36.
- Jung, J., H. Han, S. H. Ryu and W. Kim (2010). "Microsatellite variation in the pinewood nematode, *Bursaphelenchus xylophilus* (Steiner and Buhner) Nickle in South Korea." Genes & Genomics **32**(2): 151-158.
- Karaiskou, N. and C. Primmer (2008). "PCR multiplexing for maximising genetic analyses with limited DNA samples: an example in the collared flycatcher, *Ficedula albicollis*." Annales Zoologici Fennici **45**(6): 478-482.
- Kelager, A., J. Pedersen and H. Bruun (2013). "Multiple introductions and no loss of genetic diversity: invasion history of Japanese Rose, *Rosa rugosa*, in Europe." Biological Invasions **15**(5): 1125-1141.
- Keller, S. R. and D. R. Taylor (2008). "History, chance and adaptation during biological invasion : separating stochastic phenotypic evolution from response to selection." Ecology Letters **11**: 852 - 866.
- Keller, S. R. and D. R. Taylor (2010). "Genomic admixture increases fitness during a biological invasion." Journal of Evolutionary Biology **23**(8): 1720-1731.
- Kiritani, K. and N. Morimoto (2004). "Invasive insect and nematode pests from North America " Global Environmental Research **8**(1): 75 - 88.
- Kobayashi, F., A. Yamane and T. Ikeda (1984). "The japanese pine sawyer beetle as the vector of pine wilt disease." Annual Review of Entomology **29**: 115 - 135.
- Kolar, C. S. and D. M. Lodge (2001). "Progress in invasion biology: predicting invaders." Trends in Ecology & Evolution **16**(4): 199-204.
- Kolbe, J. J., R. E. Glor, L. R. G. Schettino, A. C. Lara, A. Larson and J. B. Losos (2004). "Genetic variation increases during biological invasion by a Cuban lizard." Nature **431**(7005): 177-181.
- Lee, C. E. (1999). "Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*." Evolution **53**(5): 1423-1434.
- Lee, C. E. (2002). "Evolutionary genetics of invasive species." Trends in Ecology & Evolution **17**(8): 386-391.
- Levin, D. A. (2003). "Ecological speciation: Lessons from invasives species." Systematic Botany **28**: 643 - 650.
- Levine, J. M. and C. M. D'Antonio (2003). "Forecasting Biological Invasions with Increasing International Trade
- Predicción de las Invasiones Biológicas con el Incremento del Comercio Internacional." Conservation Biology **17**(1): 322-326.
- Li, H. M., P. Y. Shen, P. Fu, M. S. Lin and M. Moens (2007). "Characteristics of the emergence of *Monochamus alternatus*, the vector of *Bursaphelenchus xylophilus* (Nematoda : Aphelenchoididae), from *Pinus thunbergii* logs in Nanjing, China, and of the transmission of the nematodes through feeding wounds." Nematology **9**: 807-816.

- Liebhold, A. M., W. L. Macdonald, D. Bergdahl and V. C. Maestro (1995). "Invasion by Exotic Forest Pests - a Threat to Forest Ecosystems." Forest Science **41**(2): 1-49.
- Linit, M. J. (1988). "Nematode - vector relationships in the pine wilt disease system." Journal of Nematology **20**(2): 227 - 235.
- Lodge, D. M. and K. Shrader-Frechette (2003). "Nonindigenous Species: Ecological Explanation, Environmental Ethics, and Public Policy." Conservation Biology **17**(1): 31-37.
- Lombaert, E., T. Guillemaud, J. M. Cornuet, T. Malausa, B. Facon and A. Estoup (2010). "Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird." Plos One **5**(3).
- Lopes, J. and S. Boessenkool (2010). "The use of approximate Bayesian computation in conservation genetics and its application in a case study on yellow-eyed penguins." Conservation Genetics **11**(2): 421-433.
- Lopes, J. S. and M. A. Beaumont (2010). "ABC: A useful Bayesian tool for the analysis of population data." Infection, Genetics and Evolution **10**(6): 825-832.
- Lounibos, L. P. (2002). "Invasions by insect vectors of human disease." Annual Review of Entomology **47**: 233-266.
- Luikart, G. and P. R. England (1999). "Statistical analysis of microsatellite DNA data." Trends in Ecology & Evolution **14**(7): 253 - 256.
- Mack, R. N., D. Simberloff, W. Mark Lonsdale, H. Evans, M. Clout and F. A. Bazzaz (2000). "Biotic invasions : Causes, Epidemiology, Global Consequences, and control." Ecological Applications **10**(3): 689-710.
- Malausa, T., A. Gilles, E. Meglec, H. Blanquart, S. Duthoy, C. Costedoat, V. Dubut, N. Pech, P. Castagnone-Sereno, C. Delye, N. Feau, P. Frey, P. Gauthier, T. Guillemaud, L. Hazard, V. Le Corre, B. Lung-Escarmant, P. J. G. Male, S. Ferreira and J. F. Martin (2011). "High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries." Molecular Ecology Resources **11**(4): 638-644.
- Mallez, S., C. Castagnone, M. Espada, P. Vieira, J. Eisenback, M. Harrell, M. Mota, T. Aikawa, M. Akiba, H. Kosaka, P. Castagnone-Sereno and T. Guillemaud (2014). "Worldwide invasion routes of the pinewood nematode: What can we infer from population genetics analyses?" Biological Invasions: 1-15.
- Mallez, S., C. Castagnone, M. Espada, P. Vieira, J. D. Eisenback, M. Mota, T. Guillemaud and P. Castagnone-Sereno (2013). "First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci." Plos One **8**(3): e59165.
- Mamiya, Y. (1972). "Pine wood nematode *Bursaphelenchus lignicolus* Mamiya and Kiyohara, as a causal agent of pine wilting disease." Review of Plant Protection Research **5**: 46 - 60.
- Mamiya, Y. (1975). "The life history of the pine wood nematode, *Bursaphelenchus lignicolus*." Japanese Journal of Nematology **5**: 16 - 25.
- Mamiya, Y. (1976). "Pine wilt disease caused by the pine wood nematode, *Bursaphelenchus lignicolus*, in Japan." Japan Agricultural Research Quarterly **10**(4): 206 - 211.
- Mamiya, Y. (1983). "Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*." Annual Review of Phytopathology **21**: 201-220.
- Mamiya, Y. (1988). "History of pine wilt disease in Japan." Journal of Nematology **20**(2): 219 - 226.
- Mamiya, Y. (2004) "Pine wilt disease in Japan". In The pinewood nematode, *Bursaphelenchus xylophilus*, ed. by M. Mota and P. Vieira, pp. 9 – 20.
- Mamiya, Y. and N. Enda (1972). "Transmission of *Bursaphelenchus lignicolus* (Nematoda : Aphelenchoididae) by *Monochamus alternatus* (Coleoptera : Cerambycidae)." Nematologica **18**: 159 - 162.
- McKinney, M. L. and J. L. Lockwood (1999). "Biotic homogenization: a few winners replacing many losers in the next mass extinction." Trends in Ecology & Evolution **14**(11): 450-453.
- McRae, B. H. (2006). "Isolation by resistance." Evolution **60**(8): 1551-1561.

- McRae, B. H. and P. Beier (2007). "Circuit theory predicts gene flow in plant and animal populations." Proceedings of the National Academy of Sciences of the United States of America **104**(50): 19885-19890.
- Metge, K. and W. Burgermeister (2008). Analysis of *Bursaphelenchus xylophilus* (Nematoda: Parasitaphelenchidae) provenances using ISSR and RAPD fingerprints. Pine wilt disease: a worldwide threat to forest ecosystems. M. Mota and P. Vieira: 175-186.
- Miller, N., A. Estoup, S. Toepfer, D. Bourguet, L. Lapchin, S. Derridj, K. S. Kim, P. Reynaud, L. Furlan and T. Guillemaud (2005). "Multiple transatlantic introductions of the western corn rootworm." Science **310**(5750): 992-992.
- Mitchell, C. E. and A. G. Power (2003). "Release of invasive plants from fungal and viral pathogens." Nature **421**(6923): 625-627.
- Molnar, R. I., H. Witte, I. Dinkelacker, L. Villate and R. J. Sommer (2012). "Tandem-Repeat Patterns and Mutation Rates in Microsatellites of the Nematode Model Organism *Pristionchus pacificus*." G3: Genes/Genomes/Genetics **2**(9): 1027-1034.
- Mooney, H. A. and E. E. Cleland (2001). "The evolutionary impact of invasive species." Proceedings of the National Academy of Sciences of the United States of America **98**(10): 5446-5451.
- Mota, M., H. Braasch, M. A. Bravo, A. C. Penas, W. Burgermeister, K. Metge and E. Sousa (1999). "First report of *Bursaphelenchus xylophilus* in Portugal and in Europe." Nematology **1**(7 - 8): 727 - 734.
- Naves, P. M., S. Camacho, E. M. de Sousa and J. A. Quartau (2007). "Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through feeding activity of *Monochamus galloprovincialis* (Col., Cerambycidae)." Journal of Applied Entomology **131**(1): 21-25.
- Nordborg, M. (2007). Coalescent Theory. Handbook of Statistical Genetics. B. M. Balding D, Cannings C. New-York, John Wiley & Sons, Ltd: 843 - 877.
- Norris, D. E., A. C. Shurtleff, Y. T. Touré and G. C. Lanzaro (2001). "Microsatellite DNA Polymorphism and Heterozygosity Among Field and Laboratory Populations of *Anopheles gambiae* s.s. (Diptera: Culicidae)." Journal of Medical Entomology **38**(2): 336-340.
- Nowak, C., C. Vogt, J. B. Diogo and K. Schwenk (2007). "Genetic impoverishment in laboratory cultures of the test organism *Chironomus riparius*." Environmental Toxicology and Chemistry **26**(5): 1018-1022.
- OEPP/EPPO (2012). "*Bursaphelenchus xylophilus* and its vectors: procedures for official control." OEPP/EPPO Bulletin **42**(3): 477-485.
- OEPP/EPPO (2014) PQR - EPPO database on quarantine pests. Available online. <http://www.eppo.int>
- Paetkau, D., R. Slade, M. Burden and A. Estoup (2004). "Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power." Molecular Ecology **13**(1): 55-65.
- Papura, D., C. Burban, M. van Helden, X. Giresse, B. Nusillard, T. Guillemaud and C. Kerdelhué (2012). "Microsatellite and Mitochondrial Data Provide Evidence for a Single Major Introduction for the Nearctic Leafhopper *Scaphoideus titanus* in Europe." Plos One **7**(5): e36882.
- Pascual, M., M. P. Chapuis, F. Mestres, J. Balanya, R. B. Huey, G. W. Gilchrist, L. Serra and A. Estoup (2007). "Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods." Molecular Ecology **16**: 3069 - 3083.
- Perdereau, E., A. G. Bagnères, S. Bankhead-Dronnet, S. Dupont, M. Zimmermann, E. L. Vargo and F. Dedeine (2013). "Global genetic analysis reveals the putative native source of the invasive termite, *Reticulitermes flavipes*, in France." Molecular Ecology **22**(4): 1105-1119.
- Pereira, F., C. Moreira, L. Fonseca, B. van Asch, M. Mota, I. Abrantes and A. Amorim (2013). "New Insights into the Phylogeny and Worldwide Dispersion of Two Closely Related Nematode Species, *Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus*." Plos One **8**(2).

- Phillips, N., M. Salomon, A. Custer, D. Ostrow and C. F. Baer (2009). "Spontaneous Mutational and Standing Genetic (Co)variation at Dinucleotide Microsatellites in *Caenorhabditis briggsae* and *Caenorhabditis elegans*." Molecular Biology and Evolution **26**(3): 659-669.
- Picard, D. and O. Plantard (2006). "What constitutes a population for the plant parasitic nematode *Globodera pallida* in its native area (Peru)?" International Journal for Parasitology **36**(1): 115-122.
- Pimentel, D., L. Lach, R. Zuniga and D. Morrison (2000). "Environmental and economic costs of non-indigenous species in the United States." BioScience **50**(1): 53 - 65.
- Pimentel, D., R. Zuniga and D. Morrison (2005). "Update on the environmental and economic costs associated with alien-invasive species in the United States." Ecological Economics **52**: 273 - 288.
- Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin and A. Estoup (2004). "GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection." Journal of Heredity **95**(6): 536-539.
- Plantard, O., D. Picard, S. Valette, M. Scurrah, E. Grenier and D. Mugni  ry (2008). "Origin and genetic diversity of Western European populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci." Molecular Ecology **17**(9): 2208-2218.
- Prenter, J., C. MacNeil, J. T. A. Dick and A. M. Dunn (2004). "Roles of parasites in animal invasions." Trends in Ecology & Evolution **19**(7): 385-390.
- Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of population structure using multilocus genotype data." Genetics **155**(2): 945-959.
- Puth, L. M. and D. M. Post (2005). "Studying invasion: have we missed the boat?" Ecology Letters **8**(7): 715-721.
- Py  sek, P., V. Jaro   ik, P. E. Hulme, I. K  hn, J. Wild, M. Arianoutsou, S. Bacher, F. Chiron, V. Did  ziulis, F. Essl, P. Genovesi, F. Gherardi, M. Hejda, S. Kark, P. W. Lambdon, M.-L. Desprez-Loustau, W. Nentwig, J. Pergl, K. Pobolj  saj, W. Rabitsch, A. Roques, D. B. Roy, S. Shirley, W. Solarz, M. Vil   and M. Winter (2010). "Disentangling the role of environmental and human pressures on biological invasions across Europe." Proceedings of the National Academy of Sciences **107**(27): 12157-12162.
- R Development Core Team (2012). R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria.
- Rannala, B. and J. L. Mountain (1997). "Detecting immigration by using multilocus genotypes." Proceedings of the National Academy of Sciences **94**(17): 9197-9201.
- Reed, D. H. and R. Frankham (2001). "How closely correlated are molecular and quantitative measures of genetic variation ? A meta-analysis." Evolution **55**(6): 1095-1103.
- Reed, D. H. and R. Frankham (2003). "Correlation between Fitness and Genetic Diversity." Conservation Biology **17**(1): 230-237.
- Rey, O., A. Estoup, M. Vonshak, A. Loiseau, S. Blanchet, L. Calcaterra, L. Chifflet, J.-P. Rossi, G. J. Kergoat, J. Foucaud, J. Orivel, M. Laponce, T. Schultz and B. Facon (2012). "Where do adaptive shifts occur during invasion? A multidisciplinary approach to unravelling cold adaptation in a tropical ant species invading the Mediterranean area." Ecology Letters **15**(11): 1266-1275.
- Rhymer, J. M. and D. Simberloff (1996). "Extinction by hybridization and introgression." Annual Review of Ecology and Systematics **27**(1): 83-109.
- Ribeiro, B., M. Espada, T. Vu, F. N  brega, M. Mota and I. Carrasquinho (2012). "Pine wilt disease: detection of the pinewood nematode (*Bursaphelenchus xylophilus*) as a tool for a pine breeding programme." Forest Pathology **42**(6): 521-525.
- Ricciardi, A. (2007). "Are modern biological invasions an unprecedented form of global change?" Conservation Biology **21**(2): 329-336.
- Ricciardi, A. and J. Cohen (2007). "The invasiveness of an introduced species does not predict its impact." Biological Invasions **9**(3): 309-315.

- Rice, W. R. (1990). "A Consensus Combined P-Value Test and the Family-Wide Significance of Component Tests." *Biometrics* **46**(2): 303-308.
- Richardson, D. M., P. Pyšek, M. Rejmánek, M. G. Barbour, F. D. Panetta and C. J. West (2000). "Naturalization and invasion of alien plants: concepts and definitions." *Diversity and Distributions* **6**(2): 93-107.
- Rius, M., X. Turon, V. Ordonez and M. Pascual (2012). "Tracking Invasion Histories in the Sea: Facing Complex Scenarios Using Multilocus Data." *Plos One* **7**(4).
- Robertson, L., S. Cobacho Arcos, M. Escuer, R. Santiago Merino, G. Esparrago, A. Abelleira and A. Navas (2011). "Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 (Nickle, 1970) in Spain." *Nematology* **13**(6): 755 - 757.
- Robinet, C., A. Roques, H. Y. Pan, G. F. Fang, J. R. Ye, Y. Z. Zhang and J. H. Sun (2009). "Role of human-mediated dispersal in the spread of the pinewood nematode in China." *Plos One* **4**(2).
- Rodrigues, J. M (2008) "National Eradication Programme for the Pinewood Nematode". In Pine wilt disease: a worldwide threat to forest ecosystems, ed. by M. Mota and P. Vieira, Springer Netherlands, pp. 5- 14.
- Rollins, L. A., A. P. Woolnough, A. N. Wilton, R. O. N. Sinclair and W. B. Sherwin (2009). "Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia." *Molecular Ecology* **18**(8): 1560-1573.
- Roman, J. and J. A. Darling (2007). "Paradox lost: genetic diversity and the success of aquatic invasions." *Trends in Ecology & Evolution* **22**(9): 454-464.
- Rosenberg, N. A. (2004). "DISTRUCT: a program for the graphical display of population structure." *Molecular Ecology Notes* **4**(1): 137-138.
- Rousset, F. (2008). "GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux." *Molecular Ecology Resources* **8**(1): 103-106.
- Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular Biology and Evolution* **4**(4): 406-425.
- Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, N. C. Ellstrand, D. E. McCauley, P. O'Neil, I. M. Parker, J. N. Thompson and S. G. Weller (2001). "The population biology of invasive species." *Annual Review of Ecology and Systematics* **32**: 305-332.
- Sala, O. E., F. S. Chapin, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker and D. H. Wall (2000). "Biodiversity - Global biodiversity scenarios for the year 2100." *Science* **287**(5459): 1770-1774.
- Sax, D. F., J. J. Stachowicz, J. H. Brown, J. F. Bruno, M. N. Dawson, S. D. Gaines, R. K. Grosberg, A. Hastings, R. D. Holt, M. M. Mayfield, M. I. O'Connor and W. R. Rice (2007). "Ecological and evolutionary insights from species invasions." *Trends in Ecology & Evolution* **22**(9): 465-471.
- Schug, M. D., T. F. C. Mackay and C. F. Aquadro (1997). "Low mutation rates of microsatellite loci in *Drosophila melanogaster*." *Nature Genetics* **15**: 99 - 102.
- Selkoe, K. A. and R. J. Toonen (2006). "Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers." *Ecology Letters* **9**(5): 615-629.
- Shibata, E. (1986). "Dispersal movement of the adult Japanese pine sawyer, *Monochamus alternatus* Hope (Coleoptera, Cerambycidae) in a young pine forest." *Applied Entomology and Zoology* **21**(1): 184-186.
- Shigesada, N., K. Kawasaki and Y. Takeda (1995). "Modeling Stratified Diffusion in Biological Invasions." *American Naturalist* **146**(2): 229-251.
- Schlötterer, C. (2000). "Evolutionary dynamics of microsatellite DNA." *Chromosoma* **109**(6): 365-371.
- Shull, G. H. (1948). "What is "Heterosis"?" *Genetics* **33**(5): 439-446.

- Siegel, S. and N. J. Castellan (1988). "Nonparametric statistics for the behavioral sciences". Statistic Series. (2nd edn.) McGraw-Hill, New York. p213-214.
- Simberloff, D. (2003). "Global climate change and introduced species in United States forests." The Science of the Total Environment **262**: 253 - 261.
- Simberloff, D. (2009). "We can eliminate invasions or live with them. Successful management projects." Biological Invasions **11**(1): 149-157.
- Simberloff, D., J.-L. Martin, P. Genovesi, V. Maris, D. A. Wardle, J. Aronson, F. Courchamp, B. Galil, E. García-Berthou, M. Pascal, P. Pyaek, R. Sousa, E. Tabacchi and M. Vilà (2013). "Impacts of biological invasions: what's what and the way forward." Trends in ecology & evolution **28**(1): 58-66.
- Sokal, R. R. and F. J. Rohlf (1995). Biometry. The principles and practice of statistics in biological research.
- Soliman, T., M. C. M. Mourits, W. van der Werf, G. M. Hengeveld, C. Robinet and A. Lansink (2012). "Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe." Plos ONE **7**(9).
- Soulé, M. E. (1992). "The social and public health implications of global warming and the onslaught of alien species." Journal of Wilderness Medicine **3**(2): 118-127.
- Sousa, E., M. A. Bravo, J. Pires, P. Naves, A. C. Penas, L. Bonifacio and M. M. Mota (2001). "*Bursaphelenchus xylophilus* (Nematoda; Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera; Cerambycidae) in Portugal." Nematology **3**: 89-91.
- Sousa, E., P. Naves, L. Bonifacio, M. A. Bravo, A. C. Penas, J. Pires and M. Serrao (2002). "Preliminary survey for insects associated with *Bursaphelenchus xylophilus* in Portugal." OEPP/EPPO Bulletin **32**: 499 - 502.
- Steiner, G., E. M. Buhner (1934) "*Aphelenchoides xylophilus*, N. SP. A nematode associated with blue-stain and other fungi in timber". Journal of Agricultural Research **48**:949 - 951
- Stocks, M., M. Siol, M. Lascoux and S. De Mita (2014). "Amount of Information Needed for Model Choice in Approximate Bayesian Computation." Plos One **9**(6).
- Stohler, R. A., J. Curtis and D. J. Minchella (2004). "A comparison of microsatellite polymorphism and heterozygosity among field and laboratory populations of *Schistosoma mansoni*." International Journal for Parasitology **34**(5): 595-601.
- Strong, D. R. and R. W. Pemberton (2000). "Biological Control of Invading Species--Risk and Reform." Science **288**(5473): 1969-1970.
- Sunnaker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll and C. Dessimoz (2013). "Approximate Bayesian Computation." PLoS Comput Biol **9**(1): e1002803.
- Suzuki, K. (2002). "Pine wilt disease - a threat to pine forest in Europe." Dendrobiology **48**: 71 - 74.
- Tares, S., P. Abad, N. Bruguier and G. de Guiran (1992). "Identification and evidence for relationships among geographical isolates of *Bursaphelenchus spp.* (pinewood nematode) using homologous DNA probes." Heredity **68**(2): 157-164.
- Tatem, A. J. and S. I. Hay (2007). "Climatic similarity and biological exchange in the worldwide airline transportation network." Proceedings of the Royal Society B-Biological Sciences **274**(1617): 1489-1496.
- Tatem, A. J., S. I. Hay and D. J. Rogers (2006). "Global traffic and disease vector dispersal." Proceedings of the National Academy of Sciences of the United States of America **103**(16): 6242-6247.
- Togashi, K. (1985). "Transmission curves of *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) from its vector, *Monochamus alternatus* (Coleoptera: Cerambycidae), to pine trees with reference to population performance." Applied Entomology and Zoology **20**(3): 246 - 251.
- Togashi, K. (1990). "A field experiment on dispersal of newly emerged adults of *Monochamus alternatus* (Coleoptera: Cerambycidae)." Researches on Population Ecology **32**(1): 1-13.
- Torchin, M. E., K. D. Lafferty, A. P. Dobson, V. J. McKenzie and A. M. Kuris (2003). "Introduced species and their missing parasites." Nature **421**(6923): 628-630.

- Tsutsui, N. D., A. V. Suarez, D. A. Holway and T. J. Case (2000). "Reduced genetic variation and the success of an invasive species." Proceedings of the National Academy of Sciences of the United States of America **97**(11): 5948-5953.
- Valadas, V., P. Barbosa, M. Espada, S. Oliveira and M. Mota (2012). "The pinewood nematode, *Bursaphelenchus xylophilus*, in Portugal : possible introductions and spread routes of a serious biological invasion revealed by molecular methods." Nematology **00**(0): 1 - 13.
- Valadas, V., M. Laranjo, M. Mota and S. Oliveira (2012). "Molecular characterization of Portuguese populations of the pinewood nematode *Bursaphelenchus xylophilus* using cytochrome b and cellulase genes." Journal of Helminthology: 1-10.
- van Wilgen, B. W., V. C. Moran and J. H. Hoffmann (2013). "Some perspectives on the risks and benefits of biological control of invasive alien plants in the management of natural ecosystems." Environmental Management **52**(3): 531-540.
- Verity, R. and R. A. Nichols (2014). "What is genetic differentiation, and how should we measure it—GST, D, neither or both?" Molecular Ecology **23**(17): 4216-4225.
- Vicente, C., M. Espada, P. Vieira and M. Mota (2011). "Pine Wilt Disease : a threat to European forestry." European Journal of Plant Pathology **133**: 89 - 99.
- Vieira, P., W. Burgermeister, M. Mota, K. Metge and G. Silva (2007). "Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses." Journal of Nematology **39**(2): 118-126.
- Viglierchio, D. R. and R. V. Schmitt (1983). "On the methodology of nematode extraction from field samples : Baermann funnel modifications." Journal of Nematology **15**(3): 438 - 444.
- Villate, L., D. Esmanjaud, M. Van Helden, S. Stoeckel and O. Plantard (2010). "Genetic signature of amphimixis allows for the detection and fine scale localization of sexual reproduction events in a mainly parthenogenetic nematode." Molecular Ecology **19**: 856 - 873.
- Vincent, B., V. Altemayer, G. Roux-Morabito, P. Naves, E. Sousa and F. Lieutier (2008). "Competitive interaction between *Bursaphelenchus xylophilus* and the closely related species *Bursaphelenchus mucronatus*." Nematology **10**: 219-230.
- Vitousek, P. M., C. M. Dantonio, L. L. Loope and R. Westbrooks (1996). "Biological invasions as global environmental change." American Scientist **84**(5): 468-478.
- Walker, B. H. and W. Steffen (1997) "An overview of the implications of global change for natural and managed terrestrial ecosystems." Conservation Ecology [online].
- Wan, X. W., Y. H. Liu and B. Zhang (2012). "Invasion History of the Oriental Fruit Fly, *Bactrocera dorsalis*, in the Pacific-Asia Region: Two Main Invasion Routes." Plos One **7**(5).
- Wang, C. Y., Z. M. Fang, Z. Wang, D. L. Zhang, L. J. Gu, M. R. Lee, L. Liu and C. K. Sung (2011). "Biological control of the pinewood nematode *Bursaphelenchus xylophilus* by application of the endoparasitic fungus *Esteya vermicola*." BioControl **56**(1): 91-100.
- Wares, J. P., A. R. Hughes and R. K. Grosberg (2005). "Mechanisms that drive evolutionary change. Insights from species introductions and invasions."
- Warren, J. E. and M. J. Linit (1993). "Effect of *Monochamus carolinensis* on the Life History of the Pinewood Nematode, *Bursaphelenchus xylophilus* " Journal of Nematology **25**(4): 703 - 709.
- Weir, B. S. and C. C. Cockerham (1984). "Estimating F-Statistics for the analysis of population structure." Evolution **38**(6): 1358-1370.
- Westley, P. A. H. (2011). "What invasive species reveal about the rate and form of contemporary phenotypic change in nature." The American Naturalist **177**: 496 - 509.
- Westphal, M. I., M. Browne, K. MacKinnon and I. Noble (2008). "The link between international trade and the global distribution of invasive alien species." Biological Invasions **10**: 391 - 398.
- Whitlock, M. C. (2005). "Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach." Journal of Evolutionary Biology **18**(5): 1368-1373.
- Whitlock, M. C. (2011). "G'st and D do not replace FST." Molecular Ecology **20**(6): 1083-1091.

- Whitney, K. D., J. R. Ahern and L. G. Campbell (2009). "Hybridization-prone plant families do not generate more invasive species." Biological Invasions **11**: 1205 - 1215.
- Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips and E. Losos (1998). "Quantifying threats to imperiled species in the United States." Bioscience **48**(8): 607-615.
- Willi, Y., J. Van Buskirk and A. A. Hoffmann (2006). Limits to the adaptive potential of small populations. Annual Review of Ecology Evolution and Systematics. **37**: 433-458.
- Williamson, M. (1999). "Invasions." Ecography **22**(1): 5-12.
- Williamson, M. and A. Fitter (1996). "The varying success of invaders." Ecology **77**(6): 1661-1666.
- Wingfield, M. J., R. A. Blanchette, T. H. Nicholls and K. Robbins (1982). "The pine wood nematode: a comparison of the situation in the United States and Japan." Canadian Journal of Forest Research **12**(1): 71-75.
- Wright, S. (1931). "Evolution in mendelian populations." Genetics **16**: 97 - 159.
- Yu, M., X. Xu and P. Ding (2011). "Economic loss versus ecological gain: the outbreaks of invaded pinewood nematode in China." Biological Invasions **13**(6): 1283-1290.
- Zhang, K., H. Liu, J. Sun, J. Liu, K. Fei, C. Zhang, M. Xu, J. Sun, X. Ma, R. Lai, Y. Wu and M. Lin (2008). "Molecular phylogeny of geographical isolates of *Bursaphelenchus xylophilus*: implications on the origin and spread of this species in China and worldwide." Journal of Nematology **40**(2): 127 - 137.
- Zhou, Z. H., D. Sakaue, B. Y. Wu and T. Hogetsu (2007). "Genetic structure of populations of the pinewood nematode *Bursaphelenchus xylophilus*, the pathogen of pine wilt disease, between and within pine forests." Phytopathology **97**(3): 304-310.